Chapter 1

Introduction

Description of:

- Motivation to investigate nearby young open clusters.

- The importance of the IMF

- Current state of knowledge in dynamical simulations.

- The problematic of constraints in dynamical theories.

It must be clear which is the objective

Description of the Nearby open clusters and their properties.

1.1 The DANCe project

- List of open clusters in the DANce project.

- the importance of the pleiades, why we restrict to it.

It must be clear what are the limitations, the boundaries in which the objective will be searched

Description of the current methodologies used to address the question mentioned previously.

- The works of Sarro, Krone-Martins, Malo, Gagne etc. LAcweing

-The advantages and caveats of the previous methodologies.

-It must be clear the necessity of a new perspective

The proposal we made. The use of Bayesian Hierarchical Models. Benefits and issues of BHM.

Description of the advantages of BHM.

-¿ It must be clear that BHM are the best choice.

Description of the practical issues needed to be solved in order to use BHM.

MCMC techinques and PSO.

-¿ It must be clear that MCMC methods are the best option.

Brief descriptions of our results and how they impact our current knowledge.

-¿ It must be clear that we attained the objective: The pleiades velocity, spatial and mass distributions.

Chapter 2

The Pleiades as a benchmark

2.1   Generalities

The ancient greeks named Pleiades to a group of stars which they believe had a common origin. These stars were the seven sisters, which, together with their parents the titan Atlas and the nymph Pleione, were put in the sky by the god Zeus.

Today we call the Pleiades cluster not just to the nine stars that made up the original Pleione family but to a much larger group, which according to Bouy et al. (2015b) goes up to ~ 2100 members. This cluster is fairly close to the sun, ~ 136 pc according to Galli et al. (2017), and is also young in galactic scales, with only ~ 120 Myr (Stauffer et al. 1998). Since it is located in the solar neighbourhood it has a distinctive velocity, when compared to that of the far distant objects, of about $-16 mas/yr$ in right ascension and $20 mas/yr$ in declination. Also, it has expelled most of its cocoon gas, which gives it an almost null extinction of $A_v = 0.12$ (Guthrie 1987).

The previous properties make the pleiades the most studied cluster in the history of astronomy. In the following sections I will describe the current knowledge on the most relevant astrophysical quantities for this work: the distance, positions, velocities, luminosities and mass of the cluster. The statistical distributions of these properties comprise the objective of the present work. I will focus primarily on the statistical aspects, however, if any I will also give notice on the modelling of these distributions.

2.2   The distance to the Pleiades

2.2.1   Measuring distances

In astronomy, measuring distances is a complicated task. Techniques vary according to the distance scale that we aim to measure. The distance ladder is constructed from smaller to larger distances. The first step in that ladder is the distance to the sun. After that, the distance to the planets and then to the stars. Since this works deal only with nearby clusters, I only explain the measuring distance to these objects.

The most direct way to measure distance to nearby stars is by means of the trigonometric parallax. This is the relative angular displacement, with respect to the far distant stars, that an object suffers in the course of a year. This relative displacement is time dependent and results from the movement of the earth (thus the observer) on its orbit around the sun. The relative displacement is maximal when measurements are taken at diametrically opposed points in the earth orbit, thus when they are separated by six months. This maximal displacement is called the parallax of the object. The distance to the object is then obtained by inverting the angular distance, measured in seconds of arc. By doing so, we obtain the distance measured in parsecs. This measurement unit gets its name from parallax-second. Thus an objects at distance one parsec from the sun shows a parallax of one arc second. The further the object is, the smaller the parallax.

As any other measurement, parallaxes had uncertainties. This uncertainties usually are a proxy for the width of the parallax distribution. Since parallaxes are related to the inverse of the distance, then the vast majority of stars had parallaxes near zero. Then, given certain precision of an instrument, and a distant object, nothing prohibits that this object may have negative measurements of its parallax. The parallax distribution is a non limited continuous distribution.

When transforming parallaxes into distances we may be tempted to take an statistics of the distribution, the mean for example, and just invert it to obtain the distance. Since this is the definition it will hold only if we have the true distance. The true distance is that in which the uncertainties are negligible. However, because measurements always have uncertainties, the inversion of the parallax will render an unbiased estimate of the true distance only for small values of the rel-

ative uncertainty (Lutz & Kelker 1973, mention that a reasonable value is below 0.15-0.20). The shape of the parallax distribution plays an important roll. If we are interested in the distance and we only have the parallax distribution, this distribution must be transformed into that of distances. However, this transformation is not a simple inversion. Several authors have proposed different approaches to the problem of distance determination using parallaxes, see for example Lutz & Kelker (1973); Bailer-Jones (2015); Astraatmadja & Bailer-Jones (2016a,b). The proper way, as Bailer-Jones (2015) points out is to infer the true distances given the observed parallaxes. For that, a prior on the distance must be established. The afore mentioned authors describe three different kinds of priors and the methodology needed to infer the true distances.

Now, I focus on the particular case of the distance to the Pleiades. The first parallax measurement of the Pleiades distance was done by van Leeuwen (1999) using the *Hipparcos* data. Later himself (van Leeuwen 2009) refined its sample and obtained a value of $120 \pm 1.9 pc$. However, Gatewood et al. (2000); Soderblom et al. (2005) using also the parallaxes of smaller samples (seven and three, respectively) of stars, measured values of $130.9 \pm 7.4 pc$ and $134.6 \pm 3.1 pc$, respectively. Finally, Melis et al. (2014) measured $136.2 \pm 1.2 pc$ using parallaxes of three stars. There is a clear controversy between *Hipparcos* data and that of the rest of the parallax measurements. The current *Tycho-Gaia* data release, TGAS, shows that the Pleiades parallax is $\pi = 7.48 \pm 0.03 \ mas$ (Gaia Collaboration et al. 2017). This seems to indicate that the *Hipparcos* parallaxes were somehow biased. However, this controversy will remain until the totally independent *Gaia* DR2.

Until this controversy have been solved, I have decided to choose the distance found by our research group, $134.4^{+2.9}_{-2.8} pc$ ($7.44 \pm 0.08 \ mas$) (Galli et al. 2017), which is in good agreement with the one found by TGAS. We found this distance using the kinematic parallaxes delivered by the moving cluster technique. This essentially exploits the fact that since clusters are bound, their members show a clear kinematic footprint: they seem to converge to a point in the sky (Blaauw 1964). Using this point and the velocity of the members (proper motion and radial velocities) it is possible to derive individual parallaxes. Furthermore, these individual parallaxes show a distribution which results from the dispersion of the cluster members along the line of sight (reference to Phillip plots). However, this

distribution is only the depth component of the space distribution of the cluster, the other two components are given by the spatial distribution.

2.3 Spatial Distribution

The spatial distribution, as I mentioned, is the two dimensional projection of the space distribution of the cluster in the plane perpendicular to the line of sight. In general, individual object positions in the plane of the sky, commonly known as the coordinates right ascension (R.A.) and declination (Dec.), are more easily measured than individual parallaxes. For this reason, just a small fraction of objects has parallaxes. In the case of the Pleiades, only ~ 70 members out of ~ 2100 have *Hipparcos*'s parallaxes. I found this figure after cross matching the Tycho-2 candidate members of Bouy et al. (2015b) with the *Hipparcos* catalogue (Perryman et al. 1997) . In addition, the relative uncertainties in R.A. and Dec. coordinates are far better (10^{-5}) than those of individual parallaxes (10^{-1}).

Due to the previous considerations, the space distribution of the Pleiades has been studied mainly trough its spatial distribution. It has been the subjects of several studies. One of the earliest results of the Pleiades spatial distribution was done by Limber (1962) who fitted the spatial distribution of the 246 candidate members of Trumpler (1921). These members were contained in a 3° radius around *Alcyone* (one of the central most massive stars of the cluster). He used a mixture of four indices polytropic distributions as described is his earlier Limber (1961) work. Later, Pinfield et al. (1998) fitted King profiles (King 1962) of different masses $(5.2, 1.65, 0.83, 0.3 \ M_\odot)$ to candidate members from the literature extending until a 3° radius. They estimated a tidal radius of $13.1pc$, in which 1194 members were contained. These amounted to a total mass of 735 $M_\odot$. The mean mass they estimated is 0.616 $M_\odot$. On the same year Raboud & Mermilliod (1998a) fitted a King's profile to a list of 270 candidate members contained within a 5° radius. They found a core radius of 1.5 pc and a tidal radius of 17.5 pc. Using different approaches their derived total mass is in the range 500 to 8000 $M_\odot$. They also measured an ellipticity of 0.17, however they do not made any explicit mention on the position angle of the axis of the ellipse.

Later, Adams et al. (2001b) also fitted a King's profile to a list of ~ 4233 objects

within a radius of 10°. These objects had membership probabilities, $p > 0.1$. They found a core radius of $2.35 - 3.0 \ \ pc$ and a tidal radius of $13.6 - 16 \ \ pc$. They estimate a total mass of $\sim 800 \ \ M_\odot$, and ellipticities in the range $0.1 - 0.35$. Converse & Stahler (2008) used a sample of 1245 from Stauffer et al. (2007) to fit a King's profile. They obtained a tidal radius of $18 \ \ pc$ and a core radius of $1.3 \ \ pc$. Later, Converse & Stahler (2010) refined their study and obtained a core radius of $2.0 \pm 0.1 \ \ pc$, a tidal radius of $19.5 \pm 1.0 \ \ pc$ and a total mass of $870 \pm 35 M_\odot$.

The previous summary of results shows at least two interesting points. The first one, is that the King's profile (King 1962) has bee the preferred choice. Although this profile has its origins on the globular clusters domain, in which the end of the cluster is well determined. It has been also widely applied to open clusters, particularly the pleiades one. The second one concern the trend of the tidal radius as a function of year of publication and size of the survey. Except the work of Adams et al. (2001b) in which the sample is large and contains low membership probability objects. Although these author mention that their high membership probability objects (1200) are contained in a 6° radius. These two aspects are bonded. The King's tidal radius will continue to increase until the physical end of the cluster will be within the survey's radius, just as it happens with the globular clusters.

2.4 Velocity Distribution

The three dimensional velocity distribution has been also studied using its projections, one along the line of sight, which corresponds to the radial velocity, and another one in the plane of the sky, which corresponds to the proper motions. These later ones are angular velocities obtained after measuring the angular displacement that an object shows when measured in at least in two different epochs. Again, measuring the stellar positions and its displacement on time is easier to obtain than the measuring of the radial velocities. These last ones are estimated through the Doppler shifting of the absorption lines in spectre of the object. Although more precise (usually on the $1 \ \ km \cdot s^{-1}$ regime) radial velocities demand large telescopes and longer observing times. For these reasons, historically the velocity distribution has been studied through the spatial velocity or proper motions distribution.

Probably the first description of the spatial velocity distribution of the Pleiades is that of Pritchard (1884). Using archival data from Knigsberg (1838-1841), Paris(1874) and Oxford (1878-1880), together with his own *Differential Microme-ter* observations, he was able to observe the relative displacements of 40 pleiades stars. According to him (Pritchard 1884): *the relative displacements of these dis-tant suns, although not distinctly and accurately measurable in numerical extent, appear to vary both in direction and amount; indicating thereby the mutual influ-ence of a group of gravitating bodies.* Later, proper motion measurements continues to be done until Trumpler (1921) use them to find the members of the cluster. He classified objects as members according to the distance they show to the mean proper motion of the cluster. This mean was calculated by Boss in his *Preliminary General Catalog* (Trumpler 1921). So far as my historic research went, this was the first measurement of an statistic of the spatial velocity distribution of the Pleiades. Later Titus (1938) using Trumpler's data an archival compilations was able to mea-sure an internal dispersion in the proper motions of 0.79 $mas \cdot yr^{-1}$. From this value he derived a mass of 260 $M_\odot$. Probably this was the first measurement of the second moment of the spatial velocity distribution.

In recent years Pinfield et al. (1998) used the velocity dispersion of the clus-ter to probe that it is in near virial equilibrium. Later, Loktin (2006) used the projected radial and tangential velocity components of the spatial velocity distri-bution of 340 members to claim the absence of evidence for rotation, expansion or compression of the cluster. Also, he does not found evidence of mass segregation in the spatial distribution. Finally, in Galli et al. (2017) we found a velocity dis-persion of 0.93 $km \cdot s^{-1}$, and the projected velocity distribution in the direction perpendicular to the great circle that joins the star with the convergent point of the cluster. This distribution (shown on Figure 10 of the mentioned work) has a dispersion of 1.45 $mas \cdot yr^{-1}$.

Concerning the radial velocities, the first record in the Pleiades is that of Adams (1904). He measured the radial velocities of six of the most brightest stars of the cluster. The latest one is the compilation of literature measurements made by Galli et al. (2017). This list contains measurements for 394 objects. This distribution is centred at 5.6 $km \cdot s^{-1}$ and is almost gaussian (see Fig. 5 of the mentioned work).

As I will explain later, the complete velocity distribution is a key ingredient

in the understanding of the cluster dynamics. Although, spatial and radial velocities are useful projections of the complete distribution, the dynamical analysis of the cluster demands the complete velocity distribution. In Galli et al. (2017) we provide a list of 64 cluster members with full spatial velocities.

2.5 Luminosity Distribution

The study of the distribution of luminosities in the Pleiades started few years later than that of the positions and proper motions. The first record I found on the luminosity distribution is the one of Trumpler (1921). He computed the number of stars in each magnitude bin for his two samples of candidate members, those comprising the objects within the central 1°, and those within 1° and 3°, referred as Tables I and II, respectively. The completeness of inner sample was estimated at 14.5 magnitudes whereas that of the outer at 9.0 magnitudes, both for the visual band. He observed that the luminosity distributions of these two samples were not alike. The inner sample is brighter than the outer one. He also observed that the luminosity distribution is not smooth and shows a local minimum at $9 - 10$ magnitudes, then an abrupt rise. Both effects in the two samples.

Figure 2.1: Luminosity distribution of Trumpler (1921)

Later, Johnson & Mitchell (1958) obtained the luminosity distribution using a sample of 289 candidate members. He assessed membership solely on photometry. Its luminosity distribution is shown in Fig. **??** Later, Limber (1962) compare the luminosity functions derived from the data of Trumpler (1921), Hertzsprung (1947), and Johnson & Mitchell (1958), see Fig. **??**These last ones are complete until absolute visual magnitudes of 8.5 and 9.5 mag. He also compared them with the initial luminosity distribution and with the present day luminosity distribution of the solar neighbourhood. He notes that the differences between observed and predicted luminosities start to happen at absolute magnitude 5.5.

In recent years, the luminosity distribution has been described in the works of Lodieu et al. (2012) and Bouy et al. (2015b). Lodieu et al. (2012) using the

Figure 2.2: Luminosity distribution in the visual band.

UKIDSS DR9 survey for galactic clusters and a probabilistic members selection method (see discussion in Section **??**) based on proper motions, and proper motions and photometry, found 8797 and 1147 candidate members, respectively. However, they do not provide the contamination rate in their analysis. Using both lists they provide their luminosity distributions in the Z band, which I show in Fig. **??**. In Bouy et al. (2015b), we estimated the present day system luminosity distribution of 1378 candidate members contained within the central 3° region (with the centre at $RA = 03 : 46 : 48$ and $Dec = 24 : 10 : 17$ J2000.0). It is called systemic because it has not been corrected for unresolved systems. An unresolved system is a group of stars (e.g. binaries) that due to its compactness appears as a single object. This distribution was computed for the K_s band and is sensitive up to $K_s \sim 20 \ mag$ and complete until $K_s \sim 17 \ mag$. This luminosity distribution is shown in Fig. 2.3

Figure 2.3: Luminosity distribution in the K_s band.

2.6   Mass Distribution

In astrophysics in general, the mass distribution is a corner stone in the understanding of the star formation process and later evolution of stellar systems. Although the temporal evolution of these systems is mainly dominated by the gravitational potential, the initial conditions and an ongoing star formation process (if any), also contribute to the shape of the mass distribution. This distribution contains the fingerprints of past events of the cluster and plays a key roll in its future evolution. Indeed, it is essential to one of the objectives of modern astrophysics: the determination of the roll played by the initial conditions or the environment, in the temporal evolution of the stellar systems.

The mass distribution of the Pleiades has been largely studied. Again, the first work on the mass distribution is that of Limber (1962). Although he did not shows any graphical or tabular representation of it, he gave the luminosity distribution and the mass-luminosity ratio. Form these the mass distributions can be derived. Instead, he use them to obtain the total mass of the cluster ($760 M_\odot$,see next section).

Most probably, the first work to present the mass distribution derived from luminosity distributions and a mass-luminosity relation from theoretical models was that of Hambly & Jameson (1991). Using R and I observations from the *United Kingdom Schmidt Telescope Unit* together with the mass-luminosity relation from theoretical isochrone models of Padova group, he was able to transform his luminosity distribution into a mass distribution, see Fig. 2.4.

Figure 2.4: Mass distribution of Hambly & Jameson (1991) derived from luminosity distribution and theoretical isochron models.

The mass distribution studies from recent years are again those of Lodieu et al. (2012) and Bouy et al. (2015b). These are shown in Fig. 2.5 and 2.6. In both the luminosity distributions are transformed into mass distributions using theoretical isochrone models. Lodieu et al. (2012) used a distance of 120.2 pc, an age of 120 Myr, and the *NEXTGEN* theoretical models Baraffe et al. (1998) to derive their mass distribution. Bouy et al. (2015b) we use a distance of 136.2 pc an age of 120 Myr and the *BT-Settl* theoretical isochrone models of Allard (2014).

Both works found contrasting aspects in their discussions. In one hand Lodieu et al. (2012) found that their present day mass distribution agrees with previous studies from the literature, and is also consistent with the system field mass function of Chabrier (2005). Chabrier (2005) fitted a log-normal function to the visual luminosity distribution of the closest 8 pc field objects. On the other hand, in Bouy et al. (2015b) we found that although the Chabrier (2005) mass function match that of the Pleiades in the $0.02 - 0.6$ $M_\odot$ mass range, it predicts to many low-mass stars and brown dwarfs.

The difference between both Pleiades present day mass distributions could arise from the different samples of members, the different theoretical isochron models,

or from both of them. The different distance values used in these two works can not account for the observed differences since they introduce only a general shift in luminosity.

Concerning the differences between the two isochrone models, in Allard et al. (2013) the authors show there are clear differences between the effective temperatures delivered by the *BT-Settl* and the *NEXTGEN* model in the low-mass regime at 5 Gyr.

Concerning the differences between the lists of candidate members, in one hand Lodieu et al. (2012) do not provide (at least explicitly) any estimate of contamination rate of their samples. Furthermore, their membership methodology has some draw-backs (see Sarro et al. 2014) that may have biased their results. Therefore, the agreement they found between their present day mass distribution and the one of Chabrier (2005), which models the field mass distribution, seem to indicate at least the following options: i) the Pleiades present day mass distribution indeed follows that of the field, or ii) their samples are contaminated.

On the other hand, in Bouy et al. (2015b) we estimated a contamination rate of 7% for which we do not have evidence of being non-homogeneous. Even if this contamination were not homogeneous, its percentage would not be able to account for the observed discrepancies between the Chabrier (2005) mass function and our present day mass distribution. These go up to $30 - 40\%$ in the low-mass regime.

The previous studies show that there is still work to do in the analysis of the Pleiades mass distribution, particularly at the low-mass range where the theoretical models, both of mass function and isochrones, led to discrepancies in the present day mass distribution.

Figure 2.5: Pleiades present day mass distribution from Lodieu et al. (2012)

Figure 2.6: Pleiades present day mass distribution from Bouy et al. (2015b)

2.6.1 Total mass of the cluster

Before ending this section I present a summary of the studies that provided an estimate of the total mass of the cluster.

The first record of the cluster total mass is that of Titus (1938). Assuming virial equilibrium he estimated it to be 260 $M_\odot$. He also computed 200 $M_\odot$ using the Eddignton's mass-luminosity relation for objects brighter than 15 mag in the visual band.

Later, the following works continue to report higher masses. Woolley (1956) estimated a total mass of 337 $M_\odot$ using Hertzsprung's catalogue. Limber (1962) computed the total mass in two ways. In the first one he assumed the cluster was virialised and obtained a mass of 900 $M_\odot$. Using the luminosity function he estimated the lower limit to the total mass in 760 $M_\odot$. Jones (1970) measured 470 $M_\odot$ and 690 $M_\odot$ using the luminosity distribution and the virial theorem, respectively. van Leeuwen (1980) using the virial theorem, a mean mass of 2 $M_\odot$, and a velocity dispersion of 0.7 $km \cdot s^{-1}$ in each spatial direction, determined a total mass of 2000 $M_\odot$. Lee & Sung (1995) measured 700 $M_\odot$ using the luminosity distribution and a mass-luminosity relation. Pinfield et al. (1998) fitting a King's profile to the spatial distribution of the cluster members obtained 735 $M_\odot$. Adams et al. (2001a) counting individual masses of candidate members within 5.5^o obtained a total mass of 690 $M_\odot$. Converse & Stahler (2008) found 820 $M_\odot$ after adding the individual masses of 1245 candidate members of Stauffer et al. (2007). To obtain these masses they transformed the K and $I - K$ magnitude and colour into masses using the mass-luminosity relation given by the theoretical isochrone models of Baraffe et al. (1998). Later, Converse & Stahler (2010) they redo their analysis and found the total mass to be 870 ± 35 $M_\odot$.

2.7 The Pleiades DANCe DR2

The Pleiades DANCe DR2 contains astrometric (stellar positions and proper motions) and photometric ($ugrizYJHK_s$) measurements for 1,972,245 objects. In Table 2.1 I provide the basic statistics for these data. Also, to complement this information, in Figs. 2.7 to 2.8 I give the uncertainties of the data set. Table 2.2 contains the number of entries in the DANCe DR2 data set that have missing

entries.

Table 2.1: Pleiades DANCe DR2 properties

Figure 2.7: default

Figure 2.8: default

2.7.1 Selection of observables

As mentioned earlier the Pleiades DANCe DR2 contains the positions $R.A., Dec.,$
proper motions ($\mu_{R.A.}, \mu_{Dec.}$ and photometric $ugrizYJHK_s$ bands of almost 2 million sources on the vicinity of the Pleiades clusters. Although these 13 observables carry information valuable to discriminate cluster members from field objects, not all of them are alike. Sarro et al. (2014) use important analysis with random forest to select the observables that were the most discriminants. Their AM, RF-2, and RF-3 reference sets comprise the $\mu_{R.A.}, \mu_{Dec.}$ proper motions and photometric bands $rizYJHK_s$. From them, these authors later excluded the r band because most of the objects in their training set do not have this observable. This resulted in a training set comprising mainly brilliant objects. In the subsequent analysis Bouy et al. (2015b) used only the RF-2, which leaves aside the z band.

In this work I only selected $\mu_\alpha, \mu_\delta, i, Y, J, H, K_s$ observables since these are the ones used Bouy et al. (2015b). This selection aims to compare our new results with the previous ones (Sarro et al. 2014; Bouy et al. 2015b). However, for the analysis of the spatial distribution we later added the stellar positions ($R.A., Dec.$).

As will be described later (Section **??**), the photometry is modelled by parametric series of cubic splines. We choose the colour index $i - K_s$ (in the following

Table 2.2: Pleiades DANCe DR2 properties

CI) to be the parameter of these series. This colour allows the most one-to-one variate-covariate relation. Figures **??** and 2.9 show the two colour-magnitude diagrams K vs $i - K_s$ and K vs $Y - J$. This last one shows second best one-to-one variate-covariate relation. This one-to-one relation is crucial to avoid degeneracies, otherwise two magnitudes could be described by the same colour index and a parametric relation would not be valid. Therefore, our photometric set of observables is $i - K_s, Y, J, H$ and K_s.

Figure 2.9: K vs $Y - J$ CMD for the Pleiades candidate members of Bouy et al. (2015b)

.

2.7.2   Data preprocessing

Since both photometry and proper motions carry crucial information for the disentanglement of the cluster population, we restrict the data set to only those objects with observed proper motions, and also at least two photometric entries in our photometric set ($i - K_s, Y, J, H, K_s$). These restrictions exclude 22 candidate members of Bouy et al. (2015b). Unfortunately, these objects have only one observed value in the photometry. For these particular objects, we compute their marginalised membership probabilities once the parameters of the model were inferred. Their membership probabilities classify XXXX of them as members.

Furthermore, we restrict the lower limit of the CI to the value of the brightest cluster member, $CI = 0.8$. We do not expect to find new bluer members in the bright part of the CMDs. Also, we set the upper limit of the CI at one magnitude above the colour index of the reddest known cluster member, $CI = 8$, thus allowing for new discoveries. Due to the sensitivity limits of the DR2 survey in i and K_s bands, objects with $CI > 8$ have K_s magnitudes ≥ 16 mag. This combination of CI and K_s magnitude is incompatible with the cluster sequence. Thus, we discard

a priori these 695 objects as cluster members. With all the previous restrictions the DANCe DR2 data set was reduced to 1 424 893 objects.

In recent days, our research group has developed a GPU version of the methodology that I present here. This GPU version is about ten times faster than the old CPU one that I developed. However, here I present the results obtained using the CPU version, the fastest one available at that moment.

Our computational constraints (associated to the CPU version) and the costly computations of our methodology (see Sect. **??**), prevented its application to the entire data set. However, since the precision of our methodology, as that of any statistical analysis, increases with the number of independent observations, we find that a size of 10^5 source for our data is a reasonable compromise. Although a smaller data set produces faster results, it also renders a less precise and potentially more biased model of the field (in the area around the cluster) and therefore, a more contaminated model of the cluster. Thus, our data set was restricted to the 10^5 objects with highest membership probabilities according to Bouy et al. (2015b). The majority of these objects ($\approx 98\%$) belong to the field with cluster membership probabilities about zero, according to Sarro et al. (2014); Bouy et al. (2015b). Thus, under the assumption that the membership probabilities given by these authors are correct, the probability of leaving out a cluster member is negligible. For the remaining of the objects in the Pleiades DANCe DR2, we assign membership probabilities *a posteriori*, once the cluster model is constructed.

Chapter 3

Bayesian formalism

This chapter provides a general introduction to probability theory and its application to parametric inference. Since the objective of this work is to infer the probability distributions of the cluster properties (e.g. luminosity and velocity), I give reason to prove that the Bayes' theorem provides the proper framework for the inference of the parameters governing these distributions. Later in this chapter, I describe the reason why the Bayesian Hierarchical Models are the best option to parametric inference under the Bayesian framework.

In the following Sections I will describe in detail the assumptions I made to model the data and to select the prior distributions. The two final Sections of this Chapter focus on: the practical issues related to the sampling of the cluster distributions, and the description of details and assumptions embbeded in the codes I developed.

Partial results of the work presented here have been submitted to the Journal A&A as Olivares et al. (2017).

3.1 Introduction to probability theory.

Uncertainty and probability are closely entangled. Anything we measure has an associated uncertainty, otherwise is not a complete measurement [1]. The term uncertainty must not be confused with the term error, which refers to the difference

---

[1] Upper and lower limits are examples of incomplete measurements.

between the measured value of the quantity and the *true* value[2] of it (JCGM 2008). It is commonly agreed that uncertainty of a measurement can be expressed in a probabilistic basis (JCGM 2008). This means that whenever we measure a quantity, lets say a, the distribution of the repeated measurements of a, follows a probability distribution function, $p(a)$. As any other probability distribution, $p(a)$ satisfies the following properties:

Property 1 It has units, those of the inverse of a.

Property 2 $p(a) \geq 0$.

Property 3 $1 = \int_a p(a)da$.

These properties hold regardless of the dimension of a. This means that the joint uncertainty of all measured quantities of an object is also a probability distribution. Furthermore, they also hold for conditional probability distributions. Let's imagine that we measure the positions, projected in the plane of sky (the plane perpendicular to the line of sight), of one star. The measurements are conditioned on the magnitude, or brightness, of the object we measure. If the object is too bright, like the sun, it saturates the detector and that renders the measurement useless. On the other hand, if the object is too faint, there are not enough photons to measure to trigger the detection. So, the stellar positions in the sky, which we can call a and b because they are two dimensions, follow a probability distribution which is conditioned on the magnitude, c, of the object. Therefore, $p(a, b|c)$ must also satisfy:

- It has units of $a^{-1}b^{-1}$.

- $p(a, b|c) \geq 0$.

- $1 = \int_a \int_b p(a, b|c)da \cdot db$.

The link between joint and conditioned probabilities is given by the following symmetric definition:

[2]The true value is that which ideally results when the uncertainty tends to zero.

$$p(a, b) = p(a|b) \cdot p(b).$$
$$p(a, b) = p(b|a) \cdot p(a). \tag{3.1}$$

This can be further conditioned on c to obtain:

$$p(a, b|c) = p(a|b, c) \cdot p(b|c),$$
$$p(a, b|c) = p(b|a, c) \cdot p(a|c), \tag{3.2}$$

If the joint probability of a and b can be factorised, this is

$$p(a, b) = p(a) \cdot p(b), \tag{3.3}$$

then a and b are say to be *independent*. An alternative option is to say that a and b are *independent*, if the conditional probability of a on b is $p(a|b) = p(a)$.

Property 3 establishes that the amount[3] of probability $p(a)$ spread over the volume of the support of a adds to one. This Property allows us to *marginalise* any non-desired variable. Let's imagine that some one gave us the probability distribution $p(a, b)$, where a and b are the positions of some star in the plane of the sky. If we are interested in, let's say, the mean value of a, we first must get rid of b. For it, we *marginalise* it out in the following way,

$$p(a) = \int_b p(a, b) \cdot db. \tag{3.4}$$

Then we compute the *expected value* of a, $E(a)$, which is identified with the mean of a once we have drawn many realisations from its probability distribution. To compute it, we add all the possible values of a weighted by their probability. This is,

$$E(a) = \int_a a \cdot p(a) \cdot da. \tag{3.5}$$

---

[3]Which could be infinite, like in Dirac's delta.

Once again, these last two equations (3.4 and 3.5) hold in case they are conditioned in any other measurement. For example, the magnitude of the object, as in our previous analogy. Notice however that, once the brightness of the object lay within the dynamic range of the detector, a and b became *independent* of the magnitude.

It is important to recall that the term measurement, and its unavoidable uncertainty, refer not just to directly measured quantities, like the photons (counts) and pixels in a CCD, but also to indirect measurements. Stellar magnitudes and positions in the sky, for example, are indirect measurements derived from the direct measurement of photons, pixels and telescope arrangement. This generalisation also applies to the measurement of parameters in any physical or statistical model, like the one I will describe in the following Section.

3.1.1 Bayes theorem

The definition of conditioned probability (Eq. 3.2) leads to the Bayes' theorem:

$$p(a|b, c) = \frac{p(b|a, c) \cdot p(a|c)}{p(b|c)}. \tag{3.6}$$

Integrating on a we find that,

$$p(b|c) \cdot \int_a p(a|b, c) \cdot da = \int_a p(b|a, c) \cdot p(a|c) \cdot da$$

$$Z \equiv p(b|c) = \int_a p(b|a, c) \cdot p(a|c) \cdot da. \tag{3.7}$$

In this last equation Z refers to what is known as the *evidence*. I will come back to this *evidence* in a few paragraphs. This last Eq. also illustrates that $p(b|c)$ is a normalisation constant which can be evaluated once $p(b|a, c)$ and $p(a|c)$ are known.

3.1.1.1 Models and parametric inference

In a broad sense, models are representation or abstraction of the *a priori* knowledge some one has about something. Sometimes this knowledge is also shared by others, sometimes it is not. Models are everywhere in our daily life: from the words we speak every day, to the evolution of the species and the general relativity; from

a kid's drawing to the cosmological models. In science, however, we restrict the concept of model to a mathematical representation of the relations (the *a priori* knowledge) among the entities that the model attempts to describe. These entities include the observables, and the parameters, if the model is parametric. In this later case, the observables can be reproduced using the parameters values. Thus, a parametric model $\mathcal{M}$, can be thought of as a function which relates parameters and observables in a mathematical way.

The act of finding the parameters that reproduce the observables is called parametric inference. This last can focus either on the entire probability distribution of the parameters, or just on some statistics about it (e.g. the maximum a posteriori, also known as MAP, or the mean and variance).

The proper way to obtain the entire probability distribution of the parameters in a model, given the data, is through Bayes' theorem. This way of parametric inference is called Bayesian inference. Another way of parametric inference is for example the maximum likelihood approach. Despite that it properly obtains the parameters that reproduce the data, it does not return their probability distribution.

In this work, as mentioned earlier, the objective is to obtain the probability distribution of the parameters in our model. Thus, I chose the Bayesian approach.

In this context, Bayes' theorem is:

$$p(\theta|\mathbf{D}, M) = \frac{p(\mathbf{D}|\theta, M) \cdot p(\theta|M)}{p(\mathbf{D}|M)}. \tag{3.8}$$

where $\theta, \mathbf{D}$ and M correspond, respectively, to the parameters in the model, the data which the model tries to describe, and the the model itself.

The term on the left is called the posterior probability distribution of the parameters,θ given the data $\mathbf{D}$ and the model M. In the right side, the two terms in the numerator are called *likelihood*, $p(\mathbf{D}|\theta, M)$ and *prior*, $p(\theta|M)$. As mentioned earlier, the denominator, $p(\mathbf{D}|M)$, is called the *evidence*.

Formally, the likelihood and the prior are probability distributions on the data $\mathbf{D}$ and the parameters θ, respectively. However, for the posterior to be a probability distribution on the parameters, it only suffices that the product of the likelihood

times the prior does not vanish everywhere or be negative anywhere[4]. If these are not probability distributions, they are called *improper* priors or likelihoods. In the extreme case that their product vanishes everywhere, which may be the case if the prior is terribly specified or if the likelihood does not take proper account of extreme data, the posterior will not be a probability distribution due to a division by zero. Nevertheless, it makes no sense to try to estimate the parameters of a model with zero evidence.

When measuring the property of individual objects in a collection of them, almost always it is assumed that measurements, and therefore their probability distributions, are independent within objects of the collection. For example, if we were to measure the weight of a group of persons, we usually assume that the weight of one person, is independent of that of another person. More properly, we usually assume that the probability distribution of the weight of one person is independent of that of another person. This assumption, however, does not always holds. Imagine for example, that we were conducting a statistical analysis on the length of the objects used to define our unit of length. In this case, the statistical distribution of the length of any object depends on the unit of measurement, which in turn is depends on the distribution of measured lengths of the rest of the objects. Therefore, the individual probability distributions of the length of the objects are not independent within them[5].

Assuming that the probability distribution of the measured values, d of a collection of N objects are independent within them, then

$$p(\mathbf{D}) = \prod_{n=1}^{N} p_n(d_n), \quad n = \{1, 2, ..., N\} \tag{3.9}$$

with p_n explicitly stating that the individual probability distributions are distinct.

Thus, the joint probability $p(\mathbf{D}|\theta, M)$ can be expressed as,

---

[4]Although negative probabilities may have sense in quantum mechanics. See for example Dirac (1942)

[5]This happens for example in satellite surveys for which their system of reference is defined by their own measurements. To avoid this issue their reference systems are anchored on independent measurements.

$$p(\mathbf{D}|\theta, M) = \prod_{n=1}^{N} p_n(d_n|\theta, M), \quad n = \{1, 2, ..., N\}. \tag{3.10}$$

The term $p_n(d_n|\theta, M)$ is the likelihood of datum d_n. This is also called the *generative* model, since it contains the necessary information to generate the data[6].

I interpret Bayes' theorem, as the probabilistic way to update knowledge. To me, this relation embodies the process of knowledge improvement once we recognise that knowledge is uncertain. In my perspective, knowledge is always uncertain. Even when its uncertainty is negligible under the current evidence that supports it. Bayes' theorem helps us to update our prior knowledge once we multiply it by the likelihood of the data. Then, the posterior probabilities, became our new knowledge. Furthermore, the Bayes' theorem also provides the objective way to compare two models or hypothesis, and update the prior information, M, used to construct them. This is called model selection, which I explain briefly in the next section.

3.1.2   Model Selection

Whenever we have a data set and two or more models that attempt to describe this data, the most straightforward thing to do is to compare these models. Almost always, we want to select the *best* model. Obviously the term *best* depends on the objective of research. For example, let's imagine that our data set consists of a set of bivariate points, for example the measurements of the positions of an object as function of time. If we were interested in reproducing exactly the same points in the data set, the *best* model will be a polynomial with degree equal to the number of points. This polynomial will pass trough all these points. Once we recognise the unavoidable uncertainty of the data, we realise that an exact representation of the data is of poor use. It fits also the noise of the data.

In general, we are interested in the predictive capabilities of a model, its ability to predict future observations rather than to replicate the ones we currently have. Thus, an exact representation of the observed data (an over-fitted model as in the previous example), will poorly describe any new data set. In this sense, an

---

[6]It may also contain the uncertainty process that transform the *true* data into the noisy observations.

over-fitted model *memorises* the data rather than *learns* from them.

A model that *learns* from the data is that which obtains the *true* underlying relation embedded in the data. This *true* underlying relation produces the *true* data. The observed data results once the uncertainty process is added. Nevertheless, we still need to select among different learning models.

We can draw some help from the commonly known Ockham's razor or principle [7]. It says:

Among competing hypotheses, the one with the fewest assumptions should be selected.

Here, hypotheses can be identified with models. Thus, this principle tells us we should choose the model with the fewest assumptions. I classify the assumptions of a model in two groups: fixed and free ones. The fixed assumptions belong to what I previously described as the *a priori* knowledge used to construct the model. These may render the model more interpretable in the physical or statistical sense, or even give it coherence within the corpus of a theory. The free assumptions on the other hand, correspond directly to the parameters in the model. They give it flexibility when fitting the data[8]. For example, in the case of a straight line model, the fact that the data is linearly related can be considered as a fixed assumption. The free assumptions correspond to the slope and ordinate at the origin.

When comparing a linear model to a quadratic one in which the constant term has been fixed, we see that they have the same number of free parameters, two, but clearly the second one has an extra fixed assumption. Therefore, choosing the model with fewer free parameters does not necessarily means choosing the model with the fewest assumptions.

One of the great advantages of the Bayesian methodology is that it incorporates directly the Ockham's principle. Suppose that we want to compare two models, M_1 and M_2, which we assume describe the data set $\mathbf{D}$. Each model has prior probabilities, $p(M_k)$ and likelihoods $p(\mathbf{D}|M_k)$ (with $k = 1, 2$). Notice that now, I use Bayes' theorem for models and not for parameters within a model, as before. So, the prior probabilities of the models reflect our believes about the fixed as-

---

[7]The origin of this motto and its exact phrasing is beyond the scope of this work. I just mention that paradoxically, an ancient formulation is attributed to Ptolomey: "We consider it a good principle to explain the phenomena by the simplest hypothesis possible" (Franklin 2002)

[8]However, they can also introduce degeneracy in the parametric space.

sumptions within each model. On the other hand, the likelihood of the data given the model, is related to the parameters (the free assumptions) and their prior probabilities, both within a model. This likelihood of the data given the model corresponds to the *evidence* of the model, Eq. 3.7. This evidence in terms of the model parameters, θ_k, is now

$$p(\mathbf{D}|M_k) = \int_{\theta_k} p(\mathbf{D}|\theta_k, M_K) \cdot p(\theta_k|M_k) \cdot d\theta_k. \qquad (3.11)$$

Bayes' theorem applied to models instead of individual parameters tells us that

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k) \cdot p(M_k)}{p(\mathbf{D})}. \qquad (3.12)$$

with $k = 1, 2$. Since there are only two models, their prior probabilities are related by $p(M_1) = 1 - p(M_2)$. Therefore,

$$p(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k) \cdot p(M_k)}{p(\mathbf{D}|M_1) \cdot p(M_1) + p(\mathbf{D}|M_2) \cdot p(M_2)}. \qquad (3.13)$$

From this last Equation, the ratio of the posterior distributions is:

$$\frac{p(M_1|\mathbf{D})}{p(M_2|\mathbf{D})} = \frac{p(\mathbf{D}|M_1) \cdot p(M_1)}{p(\mathbf{D}|M_2) \cdot p(M_2)}. \qquad (3.14)$$

This ratio provides an objective measure of how better the model M_1 is when compared to model M_2, under the measure provided by the data $\mathbf{D}$ by means of the evidence. When both prior probabilities $p(M_1)$ and $p(M_2)$ are set equal, the ratio of posteriors equal the ratio of likelihoods. This is known as the *Bayes factor* (for a similar derivation and some examples of its application see Kass & Raftery 1995).

Even when the models priors are set alike, the evidences themselves, Eq. 3.11, embody the Ockham's principle. Indeed, the evidence is a measure of the priors times the likelihood, this time for parameters in a single model. The evidence must be integrated in the parametric space. Thus, the larger the number of parameters (free assumptions), the greater the volume over which the integral must be carried on. Since the likelihood is not a probability distribution on the parameters, but on the data, when integrating in the parametric space it does not add up to one.

Thus, the likelihood can be thought as a weight that multiplies the prior in each volume element. The best model is that which maximises both the prior and the likelihood at each volume element.

As explained in this section, the paramount importance of Bayes' theorem come from the fact that it is the proper probabilistic way to update knowledge based on evidence, either if this knowledge refers to models or to parameters within a model.

3.1.3   Membership probability

In the previous Section, I derived, by means of Bayes' theorem, the probability of models M_1 and M_2 given the data $\mathbf{D}$. Now, I describe the same problem but instead of the likelihood of a data set I do it for a single datum. Here, I find the probability of model M_1 or M_2, given the datum $\mathbf{d}$, or the probability that the datum $\mathbf{d}$ was generated by model M_k. This is known as the membership probability of the datum $\mathbf{d}$ to model or class, M_k ($k = 1, 2$). The Bayes' theorem in this case is,

$$p(M_k|\mathbf{d}) = \frac{p(\mathbf{d}|M_k) \cdot p(M_k)}{\sum_{k=1}^{2} p(\mathbf{d}|M_k) \cdot p(M_k)} \tag{3.15}$$

3.2   Bayesian hierarchical Models

3.2.1   Generalities

Bayesian formalism requires the establishment of priors. As mentioned before, priors represent the *a priori* believe the user of the model has about the possible values that parameters of the model can take. This is indeed subjective. This subjectivity is the main source of criticism from the non-bayesian community [9].

Bayesian hierarchical models, in the following (BHM) can be grouped into the Empirical Bayes methods. In these later ones, the prior distributions are inferred from the data, rather than being directly specified as in common Bayesian methods. In BHM, priors are specified by parametric distributions whose parameters

---

[9]See Gelman (2012) for a discussion on the ethical use of prior information

(called hyper-parameters) are also drawn from a parametric distribution in a hierachical fashion. For this reason, hierarchical models are also called multilevel models. A fully-BHM is that in which the parameters at its higher hierarchy are drawn from a non-parametric distribution. Given its properties, BHM represent the most objective way to the establishment of prior distributions (Gelman 2006). Despite the possible high hierarchy of a BHM, for it to be valid, the class of prior distribution must allow the *true* value of its parameter (Morris 1983). For this reason, the updating of knowledge is an important step in any Bayesian study. If the likelihood is not fully allowed by the prior (as in a truncated prior, for example), then we must update it and allow the data to be fully expressed. Otherwise, the posterior could be biased.

Despite its theoretical advantages, BHM are difficult to evaluate since they require far more parameters than standard Bayes methods. Furthermore, their hierarchy (levels) must stop at some point. There are at least two approaches to stop this hierarchy. The first one is to use a non parametric distribution for the hyper-parameters at the higher level. This renders, as previously said, a fully-BHM. However, to use a non-parametric distribution we must have certain prior knowledge about it, which, most of the time is not the case. Another more practical alternative is to give a point estimate, usually the mean or the mode, for the distribution of the hyper-parameter at the top of the hierarchy.

Although in BHM the parameters values of the prior distributions are inferred from data, the user of the model has the important task of specifying the kind of distribution to be used for the prior. Selecting the kind of prior distributions continues to be an active area of research. Common options include conjugate, objective, non-informative, and weakly informative priors. Conjugate priors are those in which the posterior distribution turns out to be in the same family as the prior distribution, they are called the conjugate of the likelihood. Objective priors XXXX. Non-informative and weakly informative priors, as their names indicate, provide intentionally weaker information or no information at all for the prior. Weakly informative priors are the recommended ones (see for example the works of Gelman 2006; Huang & Wand 2013; Chung et al. 2015). Despite the kind of prior distribution chosen, we must always evaluate the prior distribution in terms of the posterior, and check if this last make sense (Gelman 2006; Gelman et al.

2013, Chap. 6).

3.2.2 Examples

Since BHM usually need more parameters than standard techniques, it restricted its use until modern computers were widely available. Although the idea of BHM was already present in the 1960s, its application to inference of normal distributions and linear models appears in the 1970s (see Good 1980, for an historical perspective of BHM). In modern days, BHM have a wide range of applications. Some examples of its application are in Gelman & Hill (2007) for the social sciences, Fei-Fei & Perona (2005) for vision recognition and, Diard & Bessiere (2008) for robot navigation.

BHM are widely applied in astrophysics. Although, originally its applications were mainly in the domain of cosmological parameters inference (see for example the works of Feeney et al. 2013; March et al. 2014; Anderes et al. 2015; Shariff et al. 2016; Alsing et al. 2017), its use was addopted in other domains. Some examples include the study of: the eccentricity distribution of binary stars Hogg et al. (2010b), the Cepheids (Barnes et al. 2004) and RR Lyrae distances (Jefferys et al. 2007), the chemical composition (Wolfgang & Lopez 2015) and albedos of exoplanets (Demory 2014), extinction maps (Sale 2012), stellar parameters (Shkedy et al. 2007), and the present day mass function (Tapiador et al. 2017).

3.2.3 Graphical representation.

Due to the generally large number of parameters in BHM, its interpretation benefits from a graphical representation. Probabilistically Graphical Models (PGM) are graphs that depict the conditional relation among the elements in a model. The elements in a probabilistic model could be constants or stochastic variables. The conditional relations that link elements could be deterministic or stochastic.

In PGM, stochastic variables are represented with circles while constants with squares. If the variable is known, as in the case of the data, it is represented with a filled symbol, otherwise with an empty symbol. Stochastic relations are depicted with solid lines while deterministic ones with dashed lines. If there is no line

between two given variables, it indicates that they are assumed to be independent. Variables that repeat together, as in the case of the data, are grouped within a plate. The number of repetitions is indicated in one corner of the plate. The community generally agrees in these set of standard representations (for more details on PGM see for example the book of Koller & Friedman 2009). Figure 3.1 shows a simple example of a PGM for the inference of the parameters of a normal distribution.

Figure 3.1: PGM representing the parametric inference of a normal distribution.

3.3   Modelling the data

Creating a model is a complex task. As previously mentioned, a model is the mathematical representation of the knowledge about something. In this work my objective is the modelling of the DANCe data related to NYOC (nearby young open clusters). Since my aim is the statistical description of the NYOC population, most of the time, the relations I use are statistical.

Modelling a data set demands the gathering and sorting of the *a priori* knowledge. This refers to the knowledge about the data set, the object of study, the statistical techniques that may help to attain the objective, and the computational resources at hand. I collected this knowledge from three main sources: the standard references (e.g. articles and books), my colleagues and experts (*knowledge elicitation*), and my self. Arrange the prior knowledge in the model is an iterative, and therefore continuous process. In this Section, I describe a snapshot of this process: the state of the model once the article Olivares et al. (2017) was submitted. Later in Section 3.6 I will give a brief description of the development of the model in the context of the coded versions of it.

In this Section I will describe one crucial aspect of the DANCe DR2 data set: the missing values. Then I will describe the relevant knowledge of NYOC and how I embedded this knowledge in the the data model in terms of field and cluster models. Finally, I will describe the details of the prior distributions and of its

hierarchy.

In the following, whenever I use the pronoun *we*, it refers to the authors of Olivares et al. (2017), where a synthesis of this work is presented. I use the pronoun I to emphasis that the particular idea or task was done by me.

3.3.1   Missing values

Missing values can happen due to different processes. From the physical perspective, they can arise due to faint or bright sources that produce counts which are outside the dynamical range of the detector. They can also emerge due to for example detector or random issues (e.g. electronic failures or cosmic rays). From the statistical perspective however, the most important aspects of missing values are the probability distribution of their occurrence, and the fact that they are partially or completely missing. A partially missing value (or partially observed) is that for which an upper or lower limit is given whereas a completely missing value is simply not available at all. The DANCe survey contains only (so far) completely missing values. Upper and lower limits could also be inferred from the data provided that missing values occur outside these limits. However, I leave aside this task since missing values in DANCe data occur also at the interior of the variables domain.

In terms of probability, there is no distinction between missing values and parameters. Therefore, we can marginalise missing values as we do with any other nuisance parameter. If datum $\mathbf{d}$ has a missing entry, d_{mis}, then $\mathbf{d} = \{d_1, d_2, ..., d_{mis}, ..., d_n\}$, with n the dimension of $\mathbf{d}$. Thus, the likelihood of this datum, given model parameters θ is

$$p(\mathbf{d}|\theta) = \int_{-\infty}^{\infty} p(\{d_1, d_2, ..., d_{mis}, ..., d_n\}|\theta) \cdot dd_{mis}. \tag{3.16}$$

Throughout this work, missing values are marginalised in this way.

I made a remark of a point that may seem obvious but it is important to remember. Let $p(a)$ be a probability distribution, A a random sample of n points from it, and $p_A(a)$ the empirical probability distribution of A. Let B be a non-random sample of $p(a)$ with n elements, and empirical probability distribution $p_B(a)$. In the limit of $n \to \infty$, $p_A(a) = p(a)$ however $p_B(a) \neq p(a)$. Therefore,

$p_A(a) \neq p_B(a)$. Similarly, in data with missing values, if the missing value pattern is not random, the distribution of the completely observed data (with non-missing values) differs from that of all the data. In Olivares et al. (2017) we show this subtle but important difference in the case of the DANCe data set.

3.3.2 The generative model

Since the objective of this work is the statistical study of NYOC we must separate them from the field population. To perform this separation, we use the data, which as always, is uncertain. Therefore, the separation is also uncertain. As mentioned in Chapter 2, under the current set of variables, the cluster and the field are entangled. To probabilistically disentangle them, we must provide probabilistic models for both populations. These models are the likelihoods. With these likelihoods and prior probabilities for the cluster and field model, we are able to compute the membership probability (Eq. 3.15) of each object in our data set to the cluster. This model is assumed to represent the cluster population. First, this model must be learnt from the data.

The learning process demands a set of N binary integers $\mathbf{q}$, one q_n for each object. Each of these two possible values represent one of the two mutually exclusive possibilities: the object belongs to the cluster ($q_n = 1$) or to the field population ($q_n = 0$). Let θ and p_c be the parameters and the likelihood of the cluster. Also, ϕ and p_f the parameters and model of the field. Then, the likelihood of the data is,

$$p(\mathbf{D}|\mathbf{q}, \theta, \phi) = \prod_{n=1}^{N} p_c(\mathbf{d}_n|\theta)^{q_n} \cdot p_f(\mathbf{d}_n|\phi)^{(1-q_n)}. \tag{3.17}$$

This $\mathbf{q}$ is now marginalised using a probability for it, a prior probability or measure which is set in terms of a new and unique parameter π, which represent the *prior* probability that an object belongs to the field. Thus, the probability of $\mathbf{q}$ is

$$p(\mathbf{q}|\pi) = \prod_{n=1}^{N} (1 - \pi)^{q_n} \cdot \pi^{(1-q_n)}. \tag{3.18}$$

and the marginalisation runs as

$$p(\mathbf{D}|\pi,\theta,\phi) = \int_{\mathbf{q}} p(\mathbf{D},\mathbf{q}|\pi,\theta,\phi) \cdot d\mathbf{q}$$

$$= \int_{\mathbf{q}} p(\mathbf{D}|\mathbf{q},\pi,\theta,\phi) \cdot p(\mathbf{q}|\pi) \cdot d\mathbf{q}$$

$$= \int_{\mathbf{q}} \prod_{n=1}^{N} p_c(\mathbf{d}_n|\theta)^{q_n} \cdot p_f(\mathbf{d}_n|\phi)^{(1-q_n)} \cdot \prod_{n=1}^{N} (1-\pi)^{q_n} \cdot \pi^{(1-q_n)} \cdot d\mathbf{q}$$

$$= \int_{\mathbf{q}} \prod_{n=1}^{N} [(1-\pi) \cdot p_c(\mathbf{d}_n|\theta)]^{q_n} \cdot [\pi \cdot p_f(\mathbf{d}_n|\phi)]^{(1-q_n)} \cdot d\mathbf{q}$$

$$= \prod_{n=1}^{N} (1-\pi) \cdot p_c(\mathbf{d}_n|\theta) + \pi \cdot p_f(\mathbf{d}_n|\phi). \tag{3.19}$$

This last equality is a rather complicated derivation which can be found on Press (1997); Hogg et al. (2010a) for p_c and p_f in the exponential family. Also, a general derivation of this expression is given by Jaynes (2003). He obtains it assuming individual unknown probabilities p_n instead of q_n and marginalising over them by the aid of a prior.

Thus, the *generative model* or likelihood of the datum $\mathbf{d_n}$ is

$$p(\mathbf{d}_n|\pi,\boldsymbol{\theta}_c,\boldsymbol{\theta}_f,\mathbf{u}_n) = \pi \cdot p_f(\mathbf{d}_n|\boldsymbol{\theta}_f,\mathbf{u}_n) + (1-\pi) \cdot p_c(\mathbf{d}_n|\boldsymbol{\theta}_c,\mathbf{u}_n), \tag{3.20}$$

where $\boldsymbol{\theta}_f$ and $\boldsymbol{\theta}_c$ indicate the cluster and field parameters, while $\mathbf{u}_n$ refers to the datum uncertainty. The probabilities $p_f(\mathbf{d}_n|\boldsymbol{\theta}_f,\mathbf{u}_n)$ and $p_c(\mathbf{d}_n|\boldsymbol{\theta}_c,\mathbf{u}_n)$ are the field and cluster models, respectively. These models are explained in detail in the next two sections.

In the following, I assume that the observed quantities, even if they contain missing values, resulted from the convolution[10] of the *true* quantities with a source of uncertainty. If I were to assume that the uncertainties of individual objects were all alike, then the data set will be homoscedastic. However, this assumption is not necessary. Instead, we model the data with its intrinsic heteroscedasticity. It means

---

[10]The addition of two stochastic variables is analogous to the convolution of their probability distributions.

that individual observations have different uncertainties. However, we assume that they belong to the same family, the multivariate normal. This assumption is standard practice and is also supported by the large and heterogeneous origins of the DANCe data.

3.3.3 The field population

To model the field population, we assume that the joint probability distribution of the data can be factorised into the probability distributions of proper motion and photometry. Thus, they are assumed independent, at least conditioned on the parameters. It is also assumed that both distributions are described by Gaussian Mixture Models (GMM). The flexibility of GMM to fit a variety of probability distribution geometries make them a suitable model to describe the density of the heterogeneous data from the DANCe DR2.

A GMM is a probability distribution resulting from the linear combination of M gaussian distributions, $\pi_m \cdot \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$, with $m = 1, 2, ..., M$. Where π_m are the amplitudes or fractions, which must add to one. Thus

$$p_{GMM}(x|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{m=1}^{M} \pi_m \cdot \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \tag{3.21}$$

Notice that the number of gaussians in the mixture is not formally speaking a parameter, but rather it implies a collection of parameters. The number of these parameters increases quadratically with the number of gaussians.

According to Bouy et al. (2015b), the number of Pleiades candidate members in their data set is 2109. This means that the number of field objects dominates (99.9%) the Pleiades DR2 data set (even our restricted 10^5 data set, in which the field objects are the 98%). Thus, it can be assumed that any reclassification of candidate members will have a negligible impact on this figures. Therefore, it seems reasonable to assume that the GMM describing the field population can be fixed during the process of cluster parameters inference. Let me elaborate on this assumption.

The field objects of Bouy et al. (2015b) are those whose cluster membership probability is lower than 0.75. There are $\sim 98,000$ of these objects in our 10^5 data set. These authors selected that probability threshold based on the analysis Sarro

et al. (2014) made of their methodology when applied to synthetic data. Sarro et al. (2014) report, at probability threshold $p = 0.75$, a contamination and true positive rates of $\approx 8\%$ and $\approx 96\%$ respectively. Therefore the number of hypothetically missclassified objects,12% of the cluster members, ≈ 258, is negligible compared to the size of the data set (10^5 objects, the restricted data set). It represents a negligible fraction ($\lesssim 0.26\%$). Furthermore, under the assumption that the work of Sarro et al. (2014) is correct, the missclassified objects concentrate on cluster membership probabilities near the classification threshold. Therefore, they will also group in an area of the physical "boundary" between the cluster and the field (I call it physical because it is on the observable variables and not in the probability threshold). This is the area of highest entanglement. It does not mean that missclassified objects will not lay in the core of the cluster (objects with high membership probabilities). It means that the occurrence of these cases will be lower. This physical "boundary" will correspond, in proper motions space to a halo around the cluster centre. In photometry, however, this boundary will run all along the cluster sequence in the CMDs. All previous assumptions are there to justify that the negligible fraction of hypothetical missclassified objects is not concentrated in the physical space.

Therefore, if the missclassified objects are a few and spread over the space, their contribution to the parameters of the GMM describing the field population can be neglected. Thus, the parameters of the GMM remain fixed and out of the inference process because including them will be prohibitive

The number of gaussians in each GMM was found using the Bayesian Information criterion (BIC, Schwarz 1978). The BIC is a model selection criteria that aims at avoiding over fitting. It represents a compromise between the likelihood, $\mathcal{L}$, of the n data points, and the number of parameters, k. This is,

$$BIC = \ln n \cdot k - 2 \ln \mathcal{L}. \tag{3.22}$$

To estimate the parameters of the GMM that maximise the likelihood, we used the Expectation Maximisation (EM) algorithm. However, the missing values in the photometry prevent the use of the standard from of the algorithm (see for example Chapter 9 of Bishop 2006). Instead, the parameters were estimated with the modified version of the EM for GMM of McMichael (1996). On it, objects with missing

values also contribute to estimate the maximum-likelihood (ML) parameters. The number of gaussians suggested by the BIC for this mixture is 14.

In the case of proper motions, since they do not contain missing values, the standard EM algorithm was used for the estimation of the GMM parameters. However, the BIC favours models in which the number of gaussians is large, their fractions small and their variances also large. This behaviour was contrary to our expectations, since the distribution resembles a bivariate normal. To circumvent this issue, I added a uniform distribution to the mixture of Gaussians. I also changed the EM algorithm to take into account the modification. The BIC applied to this new mixture of distributions renders more reasonable results. This modification improves the likelihood while reduces the number of parameters. The number of gaussians suggested by the BIC for this mixture is 7, plus the uniform distribution.

Thus, the field likelihood $p_f(\mathbf{d}|\boldsymbol{\theta}_f, \mathbf{u})$ of an object with measurements $\mathbf{d}$, given parameters, $\boldsymbol{\theta}_f$, and standard uncertainties $\mathbf{u}$ is

$$
\begin{aligned}
p_f(\mathbf{d}|\boldsymbol{\theta}_f, \mathbf{u}) = [c \cdot \pi_{f,pm,0} + \\
\sum_{i=1}^{7} \pi_{f,pm,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{f,pm,i}, \boldsymbol{\Sigma}_{f,pm,i} + \mathbf{u}_{pm})] \\
\cdot \left[\sum_{i=1}^{14} \pi_{f,ph,i} \cdot \mathcal{N}(\mathbf{d}_{ph}|\boldsymbol{\mu}_{f,ph,i}, \boldsymbol{\Sigma}_{f,ph,i} + \mathbf{u}_{ph}) \right].
\end{aligned} \tag{3.23}
$$

In this equation, $\boldsymbol{\theta}_f$ refers to the set of field parameters, $\boldsymbol{\pi}_f, \boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f$, which are, respectively, the fractions, means and covariance matrices of the GMM. The first and second brackets represent the proper motion and photometric models, respectively. The first term of the proper motion model is the uniform distribution. In it, c is a constant determined by the inverse of the product of the proper motions ranges (see Table 2.1), and $\pi_{f,pm,0}$ is the fraction of this uniform distribution. The second term in the same bracket is the mixture of gaussians with means $\boldsymbol{\mu}_{f,pm}$ and covariance matrices $\boldsymbol{\Sigma}_{f,pm} + \mathbf{u}_{pm}$.

I refer the interested reader to Appendix **??**, which contains specific details of the GMM presented in this section.

3.3.4   The cluster population

Similarly to what it was assumed for the field population, for the cluster population we assumed also that its data distribution can be factorised in the product of the proper motions distributions times the photometric distribution. It is known that unresolved systems of stars (groups of stars that, given the spatial resolution of the telescope, are seen as an individual object) have an increased brightness proportional to the multiplicity of the system. In particular, if an unresolved system is made of two equally luminous objects, then its magnitude is 0.752 times brighter than that of an individual object. This is the case of an equal mass binary.

Sarro et al. (2014) show evidence of an equal-mass binaries (EMB) sequence in the Pleiades cluster. Those authors model this parallel sequence assuming that the number of objects in this sequence is 20% of the total number of members. In this work, we also model this EMB sequence. However, we do not assume its proportion and rather infer it from the data.

Unresolved multiple systems have an impact on proper motion. From an astrophysical point of view, in stellar clusters, massive objects fall towards the centre of the gravitational potential in a rate higher than that of the less massive ones. This behaviour arises from stellar encounters in which the energy exchanges results in the less massive objects gaining speed, while the massive ones lose it. From an astronomical point of view, an unresolved binary system shifts the photo centre of its images when compared to that of a single object. For the previous reasons, we model the EMB as an independent population in the proper motions, and pair this model to its photometric parallel 0.75 displaced model. This statistical model allows a comparison of the EBM population with that of the rest of the cluster. We call all non EMB objects single stars. Although this is an abuse of the terminology because there are binaries and multiple systems with different mass ratios, it keeps the text more readable.

3.3.4.1   Photometric model of equal-mass binaries and single stars

To model the cluster sequence in the CMDs, both for single and EMB, we use cubic splines for each of the $YJHK_s$ vs CI CMDs. This assumption roots on the fitting properties of splines. I tried several polynomial bases (Laguerre, Hermite,

Chebyshev) but no matter the order, they were not able to fit the sequence, particularly in the region around $CI \approx 3$, where the slope is higher. Despite their flexibility, these splines require more parameters than common polynomials. If represented in terms of B-splines series, in addition to the coefficients of the series, they require the specification of points called knots. These knots represent the starting and ending points of the spline sections. By definition, a basis spline (B-spline) of order n is a piece wise polynomial function of order n in the interval $t_0 \leq x \leq t_n$. The boundary and internal points, $\mathbf{t} = \{t_0, t_1, ..., t_n\}$, are called knots. In particular, for a given set of knots, there is only one B-spline representation of the spline, thus the name basis spline. Any spline function of order k can be represented as a series of B-splines. In particular, any cubic spline can be represented as,

$$S_3(CI, \boldsymbol{\beta}, \mathbf{t}) = \sum_i \beta_i \cdot B_{i,3}(CI, \mathbf{t}). \tag{3.24}$$

Where $B_{i,n}$ are given by the Cox-de Boor recursive formula and $\boldsymbol{\beta}$ the coefficients of the series. For more details on splines and the Cox-de Boor formula see De Boor (1978).

As mentioned in the introduction to this Section, we assume that the observed

Despite their fitting properties, B-splines have an issue when inferring simultaneously their coefficients and knots: there is multi modality in the parametric space (Lindstrom 1999). It means that at least more than one combination of parameters produces the same solution. To avoid this multi modality, I decided to keep the knots fixed throughout inference. This decision, although reduces the flexibility of the splines, allows a still better fit than common polynomials. To obtain the ML estimate of the knots I use the algorithm of Spiriti et al. (2013). This algorithm, implemented in the *freeknotsplines* R package, allows to simultaneously obtain the best truncation value for the spline series. It uses the BIC to select among different models. In order to obtain these figures, I used the candidate members of Bouy et al. (2015b). The BIC indicates that seven coefficients is the best number with knots at $\mathbf{t} = \{0.8, 3.22, 3.22, 5.17, 8.0\}$. I tested different number of knots, ranging from two to nine, with five the best configuration given by the BIC.

As I mentioned in the introduction to this Section, we assume that the observed

photometric quantities are drawn from a probability distribution resulting from the convolution of the observed uncertainties, with the *true* quantities. Here comes an expiation. I recognise that the model is far from perfect and that there are several astrophysical phenomena that it does not addresses. Either because I do not have the knowledge or because they are too complicated to model. These phenomena include but are not limited to age, metallicity and distance dispersions, unresolved systems (other than EMB), variability, transits, etc. So, instead of modelling them I assume that all of them, even the unknown, contribute to an intrinsic photometric dispersion. Given the large amount of unknown sources contributing to this intrinsic dispersion, we can safely assume that it is multivariate gaussian. Then we infer its parameters from the data. We assume that this multivariate gaussian describes both the dispersion of cluster and EMB sequences. Most importantly, we distinguish this dispersion from the photometric uncertainty of individual measurements. If we were to assume no *true* intrinsic dispersion, then any deviation from the *true* quantities should have to be explained *only* by the observational uncertainty.

In our photometric model of the cluster, photometric observations are drawn from a multivariate normal distribution with five dimensions corresponding to our photometric reference set CI, Y, J, H, K_s. The B-splines model the mean of the *true* photometric quantities, both for the cluster sequence, $\boldsymbol{t}_{ph;Cs}$, and the EMB, $\boldsymbol{t}_{ph;Bs}$. This later one displaced 0.75 into the bright side[11] of the cluster sequence. The matrix, Σ_{clus}, represents the covariance matrix of the multivariate normal distribution. It models the width of the intrinsic dispersion. Covariance matrices are symmetric and positive semi-definite. Therefore, from the 25 entries in this 5×5 matrix, only 15 are different. We use the Cholesky decomposition to obtain a triangular matrix whose parameters we infer from the data.

Thus, the mean *true* photometry is

$$\boldsymbol{t}_{ph;Cs} = \{CI, Y, J, H, K_s\}$$
$$\boldsymbol{t}_{ph;Bs} = \{CI, Y - 0.75, J - 0.75, H - 0.75, K_s - 0.75\}$$

---

[11]This magnitude displacement of 0.75 results from the addition of two equally luminous sources.

where

$$Y = \mathcal{S}_Y(CI, \beta_Y)$$
$$J = \mathcal{S}_J(CI, \beta_J)$$
$$H = \mathcal{S}_H(CI, \beta_H)$$
$$K_s = \mathcal{S}_{K_s}(CI, \beta_{K_s})$$

and β_{Y,J,H,K_s} denote the coefficients of all the splines (a 4×7 matrix), which I call $\boldsymbol{\beta}$ for the sake of simplicity.

Since the photometry of the EMB is a linear transformation, T_{Bs}, of the mean *true* photometry of cluster sequence, no extra parameters are required. Therefore,

$$\boldsymbol{t}_{ph;Cs} = \boldsymbol{S}(CI, \boldsymbol{\beta}) \tag{3.25}$$

$$\boldsymbol{t}_{ph;Bs} = T_{Bs}(\boldsymbol{S}(CI, \boldsymbol{\beta})). \tag{3.26}$$

The cluster and EMB likelihoods of an object with photometric measurements $\mathbf{d}_{ph}$ and standard uncertainties $\mathbf{u}_{ph}$ are:

$$p_{Cs}(\mathbf{d}_{ph}|CI, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) = \mathcal{N}(\mathbf{d}_{ph}|\boldsymbol{t}_{ph;Cs}, \mathbf{u}_{ph} + \Sigma_{clus}),$$

$$p_{Bs}(\mathbf{d}_{ph}|CI, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) = \mathcal{N}(\mathbf{d}_{ph}|\boldsymbol{t}_{ph;Bs}, \mathbf{u}_{ph} + \Sigma_{clus}), \tag{3.27}$$

where $\boldsymbol{t}_{ph;Cs}$ and $\boldsymbol{t}_{ph;Bs}$ are given by Equations 3.25 and 3.26, respectively.

Since the splines are parametrised by the true CI of each object, we have more parameters than objects [12]. This *true CI* is unknown even if its observed value is not missing. We solve this problem (it is a computational problem!) by marginalising these nuisance parameters (See Eq. 3.29 and 3.30). To marginalise these CIs we need a measure or prior. We provide a prior, in a hierarchical way, (thus the name Bayesian Hierarchical model) and then marginalise the parameter. This marginalisation leaves behind a precise estimate of the parameters of the prior distribution. Paradoxically, all objects, even those without a measurement of the CI contribute to this estimate. This is the force of the BHM.

---

[12] Although this sounds crazy, the rules of probability calculus do not discard this possibility.

We model the prior for the *true CI* as a truncated $(0.8 \leq CI \leq 8)$ univariate GMM with five components whose parameters are also inferred from the data. We chose five components as suggested by BIC. I applied BIC to the results found using the EM algorithm and the candidate members of Bouy et al. (2015b). I tested with larger number of components in the mixture, but the posterior distribution did not changed, thus indicating that the BIC value was a proper assumption.

This GMM is

$$p_{CI}(CI|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}) = \sum_{i=1}^{5} \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}). \tag{3.28}$$

In this last Equation, the symbol $\mathcal{N}_t$ stands for the truncated $(0.8 < CI < 8)$ univariate normal distribution.

Then, the marginalisation of CI runs as follows:

$$p_{Cs}(\mathbf{d}_{ph}|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) = \int p_{Cs}(\mathbf{d}_{ph}, CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI$$

$$= \int p_{Cs}(\mathbf{d}_{ph}|CI, \boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot p_{Cs}(CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI \tag{3.29}$$

$$p_{Bs}(\mathbf{d}_{ph}|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) = \int p_{Bs}(\mathbf{d}_{ph}, CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI$$

$$= \int p_{Bs}(\mathbf{d}_{ph}|CI, \boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot p_{Bs}(CI|\boldsymbol{\theta}_c, \mathbf{u}_{ph}) \cdot dCI. \tag{3.30}$$

In these Equations, θ_c stands for all cluster parameters related to photometry, and the first and second terms of the integrals in the last equalities correspond to Equations 3.27 and 3.28, respectively. The distribution of CI depends only on $\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}$, thus, the cluster and equal-mass binaries likelihoods of datum $\mathbf{d}_{ph}$

are

$$p_{Cs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph})$$

$$= \int \mathcal{N}(\mathbf{d}_{ph}|\boldsymbol{\mathcal{S}}(CI, \boldsymbol{\beta}), \mathbf{u}_{ph} + \Sigma_{clus})$$

$$\cdot \sum_{i=1}^{5} \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}) \cdot dCI$$

$$p_{Bs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph})$$

$$= \int \mathcal{N}(\mathbf{d}_{ph}|T_{Bs}(\boldsymbol{\mathcal{S}}(CI, \boldsymbol{\beta})), \mathbf{u}_{ph} + \Sigma_{clus})$$

$$\cdot \sum_{i=1}^{5} \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}) \cdot dCI. \tag{3.31}$$

The observed CI and magnitudes help us to reduce the computing time of the marginalisation integral. We use them to discard regions of the marginalisation integral in which the argument is almost zero (i.e. far from the measured values). Although we allow the nuisance parameter to have all its possible values, the data, through the likelihood, gives us information about the parameter distribution. To use this information, we proceed as follows: first, we compare the observed photometry to the true one (i.e. the cluster sequence given by the splines) in a grid of 300 points uniformly distributed in the domain of CI. Then, we find the point, p, from the grid which is closest to the observed photometry. To compute distance we use the Mahalanobis metric. This metric takes into account the observational uncertainty $\mathbf{u}$ and the intrinsic dispersion of the cluster sequence, Σ_{clus}. To define the limits of the marginalisation integral, we use a ball of 3.5 Mahalanobis distances around point p. Contributions outside this ball are negligible ($< 4 \times 10^{-4}$).

3.3.4.2 Proper motion model of equal-mass binaries and single stars

As mentioned before, we assume that the cluster population has two, mutually exclusive, groups: single and EMB stars. We model the proper motions of these two subpopulations with GMM, one model for each population. If the cluster is virialised (see Chapter 2), we can assume that its velocity distribution is almost Maxwellian (Maxwell-Boltzman distribution). Therefore a GMM is a reasonable

assumption. Furthermore, in the absence of external forces, a virialised system we expect to have spherical symmetry both in its spatial and velocity distributions. Thus we can safely assume that the gaussians within the GMM share the same mean, However we allow independent means for the single and EMB GMM. The assumption of spherical symmetry may be a weak one in the presence of the galactic potential. The galactic potential can perturb the cluster and deviate its spatial and velocity distribution from spherical symmetry. The ellipticity of the spatial distribution, $\epsilon = 0.17$, which has been documented by Raboud & Mermilliod (1998a) can further deviate the velocity distribution profile from spherical symmetry. Nevertheless, since we model the covariance matrix of the GMM as full covariance matrices, departures from the spherical symmetry can still be perfectly modelled by the different eigen-vectors of these matrices.

We infer the parameters of these GMMs as part of our Bayesian hierarchical model. However we a priori set the number of gaussians in each GMM. Not doing so will demand a technique in which the model parameters can be augmented. Although such techniques already exist, they are still under computational development (see Fan & Sisson 2011, for a review of reversible jump MCMC).

Using the EM for GMM and the proper motions of the candidate members of Bouy et al. (2015b) I obtained the ML estimates for the GMM likelihood. I did this for configurations of GMM ranging from one to five components. Using BIC (Eq. 3.22), I selected four and two gaussians for single and EMB GMM, respectively. Since covariance matrices are always symmetric, only three parameters are needed to fully specify the covariance matrix of these bivariate normal distributions.

Thus, the cluster (subindex Cs) and EMB (subindex Bs) likelihoods of an object with proper motions measurements $\mathbf{d}_{pm}$ and uncertainties $\mathbf{u}_{pm}$ are

$$p_{Cs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Cs}, \boldsymbol{\mu}_{Cs}, \boldsymbol{\Sigma}_{Cs}, \mathbf{u}_{pm}) = \sum_{i=1}^{4} \pi_{Cs,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{Cs}, \Sigma_{Cs,i} + \mathbf{u}_{pm})$$

$$p_{Bs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Bs}, \boldsymbol{\mu}_{Bs}, \boldsymbol{\Sigma}_{Bs}, \mathbf{u}_{pm}) = \sum_{i=1}^{2} \pi_{Bs,i} \cdot \mathcal{N}(\mathbf{d}_{pm}|\boldsymbol{\mu}_{Bs}, \Sigma_{Bs,i} + \mathbf{u}_{pm}). \qquad (3.32)$$

Finally, combining the proper motions and photometric models, the total cluster likelihood of an object with measurement $\mathbf{d}$ and uncertainties $\mathbf{u}$ is

$$
\begin{aligned}
p_c(\mathbf{d}|\boldsymbol{\theta}_c, \mathbf{u}) = {} & \pi_{CB} \cdot p_{Cs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Cs}, \boldsymbol{\mu}_{Cs}, \boldsymbol{\Sigma}_{Cs}, \mathbf{u}_{pm}) \\
& \cdot p_{Cs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}) \\
& + (1 - \pi_{CB}) \cdot p_{Bs}(\mathbf{d}_{pm}|\boldsymbol{\pi}_{Bs}, \boldsymbol{\mu}_{Bs}, \boldsymbol{\Sigma}_{Bs}, \mathbf{u}_{pm}) \\
& \cdot p_{Bs}(\mathbf{d}_{ph}|\boldsymbol{\pi}_{CI}, \boldsymbol{\mu}_{CI}, \boldsymbol{\sigma}_{CI}, \boldsymbol{\beta}, \Sigma_{clus}, \mathbf{u}_{ph}),
\end{aligned} \tag{3.33}
$$

where π_{CB} is the parameter representing the amplitude of single cluster sequence stars in the single-EMB mixture model. The photometric and proper motions likelihoods are given by Equations 3.31 and 3.32, respectively.

3.4 Priors

The Bayesian formalism is characterised by the use of priors. These priors, as mentioned earlier, represent the objective (measurable and reproducible) way to establish beliefs about the distribution of values that certain parameter may have.

In the following I describe the information used to establish both the family of the prior distribution as well as its hyperparameters. The priors we assume are intended to fall, in most of the cases, in the category of weakly informative priors. A weakly informative prior provides weaker information than the one actually available Gelman (2006). This kind of priors show better computational performance when compared to non-informative priors. Examples of these can be found in the works of Gelman et al. (2008) and Chung et al. (2015).

I group priors into three main categories. Fractions in general, and those concerning: proper motions and photometric models.

Fractions are defined for mixtures. The mixtures in the model are the GMMs, the cluster-field, and the single-EMB mixtures. At each mixture, the fractions or amplitudes quantify the contribution of each element to the mixture. If each element in the mixture is itself a probability distribution, then the fractions must add to one and be bounded by the $[0, 1]$ interval. We choose the Dirichlet distribution since it is the multivariate generalisation of the beta distribution. It has support in $[0, 1]$ for each entry, and is parametrised by $\boldsymbol{\alpha}$. This distribution gives the probability of n rival events given that each rival event has been observed $\alpha_i - 1$ times

$(i = \{1, 2, ..., n\})$.

The mean and variance of the Dirichlet distribution are given by,

$$E[x_k] = \frac{\alpha_k}{\sum_k \alpha_k}, \tag{3.34}$$

$$Var[x_k] = \frac{-\alpha_k \cdot (\alpha_k - \sum_k \alpha_k)}{(\sum_k \alpha_k)^2 \cdot (1 + \sum_k \alpha_k)}. \tag{3.35}$$

For the field-cluster mixture we set the hyper-parameters to $\boldsymbol{\alpha} = \{9.8, 0.2\}$. We expect a mean 98% of field objects and a 2% of cluster objects with little variance. These figures correspond to the fraction of field and cluster candidate members found by Bouy et al. (2015b). For the single-EMB mixture we use an hyper-parameter value, $\boldsymbol{\alpha}_{CB} = \{8, 2\}$. We expect a mean 20% of EMB, as suggested by Bouy et al. (2015b). For fractions in the proper motions GMM, hyper-parameter are $\boldsymbol{\alpha}_{Cs} = \{1, 1, 5, 5\}$ and $\boldsymbol{\alpha}_{Bs} = \{1.2, 8.8\}$. These values induce fraction distributions whose means are similar to the fractions recovered after fitting a GMM to the Bouy et al. (2015b) candidate members. For the fraction in the GMM of the CI distribution, the hyper-parameter were set all to 1, ($\boldsymbol{\alpha}_{CI} = \{1, 1, 1, 1, 1\}$), which results in equal means and large variances for all of them.

In the previous cases, with exception of the cluster-field mixture, the hyper parameters were chosen to have larger variances. Fig. 3.2 shows the distribution associated to these hyper-parameters.

The narrow variance in the cluster-field mixture expresses our prior belief about the number (fraction) of candidate members within our large data set. I believe this figure is small.

Figure 3.2: Prior distribution of fraction parameters. From top left to bottom right, the distributions of field fraction (π), equal mass binaries fraction ($1 - \pi_{CB}$), and the cluster (π_{Cs}) and equal-mass binaries (π_{Bs}) fractions in their proper motion GMM, respectively.

For the priors of mean parameters in the proper motions GMM, both of single and EMB, we choose the bivariate normal distribution. We set the hyper-parameters of this bivariate normal to the MLE found after fitting a bivariate

normal to the candidate members of Bouy et al. (2015b). These values are $\boldsymbol{\mu}_{\mu_{pm}} = \{16.30, -39.62\}$ and $\Sigma_{\mu_{pm}} = \{\{36.84, 1.18\}, \{1.18, 40.71\}\}$.

As prior for the covariance matrices of both single and EMB proper motions GMM we use the Half–$t(\nu, \mathbf{A})$ distribution. With ν an scalar and $\mathbf{A}$ a vector of dimension equal to that of the space. As shown by Huang & Wand (2013) this family distribution leads to more accurate estimation of covariance matrices than the traditional Inverse-Wishart distribution. Particularly, the marginal correlation correlation parameters, ρ have the following distribution,

$$p(\rho) \propto (1 - \rho^2)^{\frac{\nu}{2}-1} \tag{3.36}$$

while the standard deviation term σ_k associated to entry k is distributed according to Half–$t(\nu, A_k)$. We set the hyper-parameter to $\nu = 3$ and $\boldsymbol{A}_{pm} = \{10^5, 10^5\}$. According to Huang & Wand (2013), an arbitrarily large values of $\boldsymbol{A}$ lead to arbitrarily weakly informative priors on the corresponding standard deviation terms.

Photometric priors can be grouped in three categories, those concerning: (i) the *true CI*, (ii) the splines coefficients, and (iii) the cluster sequence intrinsic dispersion.

For the means in the univariate GMM modelling CI we use a uniform in the range $(0.8 \leq CI \leq 8)$. For the standard deviations I choose the Half–Cauchy$(0, \eta)$ distributions as suggested by Gelman (2006). I choose the arbitrarily large value of $\eta = 100$..

For the coefficients in the spline series we set the priors as univariate normal distributions. To find the values of the hyper-parameters, we proceed as follows. First, we remove EMB from Bouy et al. (2015b) candidate members. I performed an iterative fit of the cluster sequence, such that in each iteration objects above the 0.75 magnitude were removed. In the region of $CI > 7$ no candidate members have been found. Thus, to provide a prior we complement our list of candidate members with the brown-dwarfs from the Faherty et al. (2012) sample. Only those with the same CMD as our data set. Finally, we fit the splines, and use the coefficients of this fit as means, μ_β of the univariate normal distributions. The standard deviation terms were set to $\sigma_\beta = \{1, 1, 1, 1, 1, 0.5, 0.1\}$. These values provide a reasonable compromise between cluster sequences compatible with the previously known candidates and those far away or with exotic shapes. We show

a sample of this priors in Fig. 3.3. This Fig. also shows the brown-dwarfs from Faherty et al. (2012) and the sequence (dashed line) we use to provide the means of the univariate normal distributions.

To set the prior for the parameters of the cluster intrinsic dispersion, I choose again the Half–t$(\nu, \boldsymbol{A})$ distribution. However, this time I use $\boldsymbol{A}_{ph} = \{10, 10, 10, 10, 10\}$. These values are large when compared to the standard deviation terms o the observation uncertainty, thus provide a weakly informative prior on the marginal standard deviation terms of the Σ_{clus} covariance matrix.

Table **??** shows a summary all the hyper-parameter and their values.

Figure 3.3: CMD K_s vs. $i - K_s$ showing a sample of the prior for the coefficients in the splines series. Also shown the brown-dwarfs we add from Faherty et al. (2012) sample, and the cluster sequence (dashed line) found after fitting the splines to the brown-dwarfs and candidate members below the equal-mass binaries sequence.

3.5 Sampling the posterior distribution

Theoretically, there are at least three possible approaches to obtain the posterior distributions of the parameters in our model. One of these theoretical options is the analytical solution. This solution is obviously not feasible, due to our large data set. The second theoretical option consists of a grid in parameter space. The likelihood and the prior must be evaluated at each point in this grid and then multiplied. This approach is reasonable when the parametric space is of moderate dimension. It requires the evaluation of the posterior distribution q^p times, with q the number of grid points in one dimension, and p the dimension of the parametric space. The number of parameters in our model is 85, which immediately rules out this possibility. The third and so far only feasible approach is the use of Markov Chain Monte Carlo (MCMC). Although these methods provide a solution in a reasonable time, nevertheless, the bottle neck of computing time is due to the evaluation of the likelihood, which rows linearly with the size of the data set.

This Section is structured as follows. First, I introduce an heuristic technique to search for the maximum a posteriori of our target distribution: the posterior

distribution of the parameters given the data. Then I will describe the MCMC techniques available in the literature. Particularly the one we chose and the reasons for which it was chosen. I will end this section with the details about the assessment of the MCMC convergence.

3.5.1 PSO

The likelihood of the data is the product of the individual likelihoods of each datum (Eq. **??**). Therefore, the number of operations needed to evaluate the likelihood grows proportionally to the size of the data set. As I will explain in Section **??**, the burn-in phase of MCMC techniques allows them to reach the target distribution. However, once the MCMC reaches the target distribution, the burn-in computations are discarded. Since the evaluation of the likelihood, and therefore of the posterior, is computationally expensive due to our large data set, I decided to reduce as much as possible the burn-in phase. To do so, I provide MCMC with a set of initial solutions which are close to the Maximum-A-Posteriori (MAP) of the target (posterior) distribution. These MAP solution must not be crowded on the MAP solution, otherwise MCMC will spend a lot of time in expanding this initial set of solutions. Our objective is a representative sample of the full posterior distribution, not just an estimate of it.

In Section **??** I will show that the MCMC flavour more suitable to our objective belongs to the family of *ensemble* MCMC. This flavour works with particles in the parametric space. To make the transition between the initial MAP solutions and the particle MCMC as efficient as possible I chose a MAP finder which works also with particles. The Particle Swarm Optimiser (PSO, Kennedy & Eberhart 1995) provides a heuristic cheap and fast approach to the MAP solution. The PSO works with an ensemble of particles which move in through the parametric space. These particles use the collective and individual past and present information to update their position. This information is specified by the score function, which in our case is the posterior distribution. The particles update their position iteratively by means of a velocity. It has a random but restricted magnitude, however, its direction is determined by the object position and the individual and collective positions of maximum score. Kennedy & Eberhart (1995) has a detailed description of the original algorithm, while a more efficient version is given by (Clerc &

Kennedy 2002).

Although the PSO is a simple and rather efficient solution to the MAP approximation, it is far from perfect. Due to its heuristic origin, there is no theory behind its formulation, furthermore, it does not always guarantee the finding of the global maxima. A convergence guaranteed version can be found in Patel et al. (2013). This issue does not affect our results because MCMC does guarantee the target distribution. However, it has an impact in the computing time. Also, PSO stops once the particles scores are within a user defined tolerance. If the tolerance is too large, the PSO may not converge. If the tolerance is small it may converge but deliver solutions highly concentrated around the solution. This poses a problem to the following MCMC stage. In order to explore the full posterior distribution MCMC needs more iterations to expand the initially concentrated positions. To overcome this problem I decided to use the charged PSO (Blackwell & Bentley 2002)

3.5.1.1 The charged PSO

Originally designed to optimise a time varying score function, the charged PSO maintain its exploratory capabilities due to an electrostatic force that repels particles when they got closer than a certain distance (Blackwell & Bentley 2002). Thanks to this electrostatic force the charged PSO avoids the over-crowding of particles around local best values.

The algorithm of Blackwell & Bentley (2002) computes distances in the entire parametric space. I find this approach unsuitable to our problem, thus I modified it. This modified version is described in more details on Section 3.6.1.

3.5.2 MCMC

3.5.2.1 Generalities

Markov chain Monte Carlo is the generic name for a series of algorithms whose objective is the sampling of probability distributions. As their name indicates, MCMC generate a chain of Monte Carlo realisations that fulfil the Markov property. Monte Carlo realisations can be understood, broadly speaking as continuous random realisations. Since it is an iterative algorithm, the chain refers to the joint

of al random Monte Carlo steps. The Markov property refers to the probabilistic independence of the steps in the chain. A Markov chain is that in which the probability of a future step depends only on the present step, and not in the past steps.

Andrieu et al. (2003) provides a brief and interesting summary of the history of the MCMC methods. In the following I use their work to describe the fundaments of MCMC. For more details, see the aforementioned authors and the book of Brooks et al. (2011).

A stochastic process is defined as a sequence $\{\theta_1, ..., \theta_n\}$ of random elements from a set, where each element $\theta_i \subset \mathbb{R}^k$ with k the dimension of the *state space.*

A stochastic process, $\boldsymbol{\theta} = \{\theta_0, \theta_1, ..., \theta_n, \theta_{n+1}\}$ is called a Markov chain if

$$p(\theta_{n+1}|\theta_0, \theta_1, ...\theta_n) = p(\theta_{n+1}|\theta_n).$$

A Markov chain has two important distributions, the initial distribution and the transition distribution. The initial distribution is the marginal distribution of θ_0, it is $p(\theta_0)$. The transition distribution is the conditional probability $p(\theta_{n+1}|\theta_n)$. This last one is called stationary or homogeneous if it does not depend on n.

If this transition is irreducible and aperiodic, then there is an invariant or equilibrium distribution to which the chain converge in spite of the initial distribution. Aperiodic means that the chain does not make loops. It is irreducible if the probability of exploring all other states is not zero.

If we want to have $p(\theta)$ as the invariant distribution, then it suffices that the transition distribution $p_t(\cdot|\cdot)$ satisfies the detailed balance condition,

$$p(\theta_n) \cdot p_t(\theta_{n-1}|\theta_n) = p(\theta_{n-1}) \cdot p_t(\theta_n|\theta_{n-1}) \tag{3.37}$$

Thus MCMC are Markov chains that satisfy the detailed balance and had their invariant distribution as the target distribution. The variety of MCMC algorithms arises from the efficiencies in which they arrive to the target distribution.

In the following I will review three of the MCMC categories: Metropolis-Hasting (MH), Hamiltonian Monte Carlo (HMC), and affine invariant samplers. The MH category comprises the classic MH algorithm but also contain particular cases like the Gibbs sampler (Geman & Geman 1984). I describe MH only for completeness

and explanatory reasons. Later, I will focus on the particular cases of Hamiltonian Monte Carlo, for (HMC), affine invariant for ensemble samplers. Finally, I will briefly describe Nested Sampling, an algorithm that uses MCMC to numerically compute the Bayesian evidence while simultaneously samples the posterior distribution.

3.5.2.2   Metropolis-Hastings

By far, the most popular MCMC algorithm is Metropolis-Hastings (Metropolis et al. 1953; Hastings 1970). Given the current, θ, and proposed, $\hat{\theta}$ positions of the Markov chain, which live in the state space, the chain moves from θ to $\hat{\theta}$ with acceptance probability

$$\mathcal{A}(\hat{\theta}|\theta) = min\left\{1, \frac{p(\hat{\theta}) \cdot q(\theta|\hat{\theta})}{p(\theta) \cdot g(\hat{\theta}|\theta)}\right\}. \qquad (3.38)$$

Where q is the transition probability. Since the algorithm allows rejection, it is aperiodic, and to ensure irreducibility, the support of q must include that of p (Andrieu et al. 2003). The popularity of MH lies in its simplicity. Nevertheless it requires a careful tuning of the transition probability.

3.5.2.3   Hamiltonian Monte Calro

The Hamiltonian Monte Carlo algorithms (Duane et al. 1987; Neal 1996), as they name suggests[13], use Hamiltonian dynamics to express the target distribution as the potential distribution of a hamiltonian system. In such systems total energy is the sum of the potential and kinetic energies. The potential distribution depends only on position, whereas the kinetic one on momentum. To this end, HMC introduces a momentum to fictitious particles to use their positions as a sample of the target distribution. To update the particles positions, HMC uses the Hamilton equations. They contain information about the gradient of the potential. Once HMC has tuned the momentum distribution, the proposed positions are more likely in terms of the target distribution. Therefore, using the information about the gradient of the target distribution, HMC is able to improve the acceptance ratio

---

[13]Originally called Hybrid Monte Carlo by (Duane et al. 1987)

of the proposed steps. A detailed description of HMC can be found in Chapter 5 of Brooks et al. (2011). The package *Stan* (**?**) provides an efficient implementation of HMC.

3.5.2.4 Affine invariant

Affine invariant MCMC samplers use many particles, (ensemble), to sample the target distribution with a performance that is independent of its shape in the parametric space. Affine invariant MCMC do not need to tune the transition distribution, for this reason, these samplers are faster than standard MCMC (Goodman & Weare 2010). In the following I use the derivation of Goodman & Weare (2010).

An ensemble $\boldsymbol{\theta}$ is a set of L particles $\theta_l \subset \mathbb{R}^k$ living in state space $\mathbb{R}^{kL}$. These particles are independently drawn from the target distribution π. Therefore,

$$\Pi(\boldsymbol{\theta}) = \pi(\theta_1) \cdot \pi(\theta_2)...\pi(\theta_L).$$

Thus, an ensemble MCMC is a Markov chain in the state space of ensembles, more properly, in the state space of the sequence $\boldsymbol{\theta}(1), \boldsymbol{\theta}_(2), ...\boldsymbol{\theta}(t)$. An ensemble MCMC preserves the equilibrium distribution without the individual particles sequence, $\theta_1(1), \theta_1(2), ...\theta_1(t)$, being Markov or even independent.

To update the particles positions, the detailed balance must be fulfilled. Goodman & Weare (2010) use partial resampling to ensure this. The transition preserves the target distribution if the single particle move preserves the conditional distribution of the particle given the complementary ensemble (the rest of the particles). Using the affine invariant *stretch move*, they are able to define a Markov chain (in the state space of ensembles) that satisfies the detailed balance. The stretch move $\theta_k(t) \to \hat{\theta}$, defined as,

$$\hat{\theta} = \theta_j(t) + z \cdot (\theta_k(t) - \theta_j(t)),$$

where $\theta_j(t)$ is the current position of a particle in the complementary ensemble, and z is the stretching factor. It produces a symmetric transition, $p(\theta_k(t) \to \hat{\theta}) = p(\theta_k(t) \leftarrow \hat{\theta})$, if its density $g(z)$ satisfies the symmetry condition

$$g(\frac{1}{z}) = z \cdot g(z).$$

Finally, Goodman & Weare (2010) define their affine invariant MCMC using $g(z) \propto 1/z, z \in [1/a, a]$ and zero otherwise, together with acceptance probability,

$$\mathcal{A}(\hat{\theta}|\theta) = min\left\{1, z^{n-1} \cdot \frac{p(\hat{\theta})}{p(\theta)}\right\}. \tag{3.39}$$

The parameter $a > 1$ improve the performance the performance of the sampler (Goodman & Weare 2010).

One of the greatest advantages of ensemble samplers is its possibility of parallelisation. Since they work with particles, this particles can be distributed among cores in a computer cluster, therefore reducing considerably the computing time. Foreman-Mackey et al. (2013) implemented the affine invariant stretch move of Goodman & Weare (2010) in the Python package *emcee*.

3.5.2.5 Nested sampling

Nested sampling (Skilling 2004, 2006) is an algorithm designed to numerically integrate the evidence (Eq. 3.11). As a by product it delivers also a sample of the posterior distribution. It samples the prior and use these particles to compute integral. The integral is the sum of the likelihood, L_i of each particle times a weight, w_i. This weight is a proxy of the volume of the prior covered by the updated position of the particle, $w_i = \exp(-(i-1)/N) - \exp(-i/N)$. Particles update their position only when their likelihood is the minimum from the group of particles. In this way,

$$z \leftarrow \sum_i w_i \cdot L_i \tag{3.40}$$

An improved version of the original algorithm was implemented in *MultiNest* (Feroz et al. 2009). This version allows the sampling and computing of evidence in multimodal posteriors.

3.5.2.6 Implementation and convergence

To sample the posterior distribution in our problem, we chose *emcee* due to the following properties: i) the affine invariance allows a faster convergence over common and skewed distributions (see Goodman & Weare 2010; Foreman-Mackey

et al. 2013, for detail), ii) the parallel computation distributes particles over nodes of a computer cluster and thus reduces considerably the computing time, and iii) it requires the hand-tuning of only two constants: the number of particles, and the parameter a of the $g(z)$ distribution. I choose a particle to parameters ratio of two, this means 170 particles. This is the minimum recommended by Foreman-Mackey et al. (2013) that allows a reasonable computing time. After trial and error, I fixed the value of parameter $a = 1.3$. As mentioned by Goodman & Weare (2010) this parameter can be tuned to improve performance of the sampler. The chosen value keeps the acceptance fraction in the range $0.2 - 0.5$ as suggested by Foreman-Mackey et al. (2013).

I use a modified version of *CosmoHammer* (Akeret et al. 2013), a front-end of *emcee*. See section 3.6.4 for more details.

As mentioned earlier, the PSO does not guarantee the finding of the global maximum of the score function. Therefore, I decide to implement an iterative approach that minimises the risk of PSO to get stuck in a local maxima. To do so I iteratively run PSO and 50 iterations of *emcee* (with the same number of particles as the PSO) until the relative difference between means of consecutive iterations was lower than 10^{-7}. The iterations of *emcee* spread the PSO solution without moving away from the maximum.

Neither scheme, PSO alone or PSO-*emcee*, guarantees to find the global maximum and their solution could be biased. However, we use them to obtain a fast estimate of the global maximum, or at least, of points in its vicinity. If the initial solution provided by this scheme is indeed biased, the final *emcee* run, during the burning phase, erases any dependance on these initial solutions. After convergence of the PSO-*emcee* scheme, we run *emcee* until convergence.

Convergence to the target distribution occurs when each parameter enters into the stationary equilibrium, or normal state. The Central Limit Theorem ensures that this state exists. See Roberts & Rosenthal (2004) for guaranteeing conditions and Goodman & Weare (2010) for *irreducibility* of the *emcee* stretch move. The stationary or normal state is reached when, in at least 95% of the iterations, the sample mean is bounded by two standard deviations of the sample, and the variance by the two standard deviation of the variance [14]; see Fig. 3.4.

---

[14] $sd(\sigma^2) = \sigma^2 \sqrt{\kappa/n + 2/(n-1)}$ with κ the kurtosis and n the sample size.

Figure 3.4: Normalised mean (left panel) and variance (right panel) of each parameter in our model, as functions of iterations. The normalisation values are the mean and variance of the ensemble of particles positions at the last iteration. Red lines show one and two sigma levels of these normalisation values.

Once all parameters have entered the equilibrium state, we stop *emcee* sampling using the criterium of Gong & Flegal (2016) [15]. We chose this criterium because it was developed for high-dimensional problems and tested on Hierarchical Bayesian Models. In this criterium, the MCMC chain stops once its "effective sample size" (ESS, the size that an independent and identically distributed sample must have to provide the same inference) is larger than a minimum sample size computed using the required accuracy, υ, for each parameter confidence interval $(1-\delta)100\%$. Our *emcee* run stops once the ESS of the ensemble of walkers is greater than the minimum sample size needed for the required accuracy $\epsilon = 0.05$ on the 68% confidence interval ($\delta = 0.32$) of each parameter.

3.6 Codes

This Section sets out the details about the code I develop to perform the computation described throughout this chapter. First, I give a brief chronological description of the model development. Later, I will describe the details on the implementation of the charged PSO, the modified *emcee*, and the GMM used to describe the field population. Finally, I will end this Section detailing the hybrid high performance computing code developed to minimise the computing time of the BHM.

The first version of the Bayesian Hierarchical Model was implemented by Ángel Berihuete in the package Stan (). It comprised a Bayesian model of the ML model of Sarro et al. (2014). The proper motions were modelled using a single mixture of gaussians. The photometry was modelled with a Chebyshev polynomial parametrised with the length along the sequence. This length was found using a principal curve analysis. Later, I included the EMB and uncertainties both in

---

[15]Implemented in the R package *mcmcse* (Flegal et al. 2016)

proper motion and in photometry, and the width of the sequence modelled as the multivariate gaussian. We realised that the principal curve analysis is not compatible with the deconvolution methodology. The principal curve analysis is driven by the elements with higher variance. We decided to parametrise the polynomials with the true colour, a nuisance parameter that was marginalised with the aid of a prior. For this prior we introduced the colour distribution. Previous to the introduction of the marginalisation of the nuisance parameters, the model worked fine on samples of a few thousands of stars. One the marginalisation was introduced the model increased the amount of computing time rendering it useless for higher data sizes. At this point we decided to port the existing code into python in order to use the parallel *emcee*. *emcee* proved to be of great use. Due to its parallelisation capabilities we were able to increase the data size from 2000 to 10,000 objects. Since the computing of the likelihood was the highest computational challenge, I developed my own routines to perform it on parallel. However, *CosmoHammer* (Akeret et al. 2013) turn out to be more efficient in distributing the loads. I ported the code into *CosmoHammer* and modified this last one to better suit our needs. Despite the Hybrid-HPC, the computing of the likelihood of 10^5 seemed unreachable. At this point I performed two tasks, the first was to strip the code of all auxiliary libraries calls, and the vectorisation of the operations. Since the parameters of the field were held fixed, the field likelihood was computed externally for each object. The code was then fed with the data, the field likelihood, and the auxiliary computations reduced at minimum. These reduction included for example, the Cholesky decompositions and matrix inversions of the uncertainty. Thus, instead of doing these computation inside the code, the code was fed with the appropriate values.

Introducing PSO and later the charged PSO reduced the computing time. At this point we were able to compute the first 10^5 run, which however did not converged in reasonable time. Once the approximation of the marginalisation integral (Eq. **??**) was introduced the computing time reduced far more. Finally, the tuning of the parameter allowed us to increase the acceptance fraction, and reach convergence within 4 weeks of full computing time in a 80 cores computer cluster. It is indeed a very long time, however it is reasonable compared with our original

estimates of approximately 2 years[16].

3.6.1   The modified charged PSO

As explained before, the charged PSO of Blackwell & Bentley (2002) was inappropriate to our objective. The metric of the parametric space of our problem is not isotropic, parameters have different length scales. For example, while fraction are constrained in the $[0, 1]$ interval proper motions parameters can go far larger than that. Therefore, the use of an isotropic metric results in a solution which is crowded in some parameters while is over dispersed in others. To solve this issue, I modified the charged PSO by measuring distance between particles and applying the electrostatic force independently in each parameter. In such a way, the electrostatic force plays a role only when the relative distance between particles is smaller than 10^{-10}. I found this value heuristically.

In the original version of Blackwell & Bentley (2002), each particle is subject to the acceleration,

$$\mathbf{a} = \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}^3} \cdot \mathbf{r}_{ij}, \qquad p_{core} < r_{ij} < p \tag{3.41}$$

where q_i and q_j are the charges of particles i and j, r_{ij} is the distance between them. The distances p_{core} and p indicate the minimum and maximum distances at which the electrostatic force came into action. Outside this range, the electrostatic force is zero. In this equation, $\mathbf{r}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, where $\mathbf{x}_i, \mathbf{x}_j$ are the positions of particles i and j. Also, $\mathbf{r}_{ij}, \mathbf{x}_i, \mathbf{x}_j \subset \mathbb{R}^d$, with d the dimension of the space.

In the modified version, the distance is measured independently in each dimension of the parametric space. Thus, $\mathbf{r}_{ij} = \{x_{1,i} - x_{1,j}, x_{2,i} - x_{2,j}, ..., x_{d,i} - x_{d,j}\}$. Also the acceleration has the form,

$$\mathbf{a} = \sum_{i \neq j} \frac{q_i \cdot q_j}{r_{ij}^2} \cdot \mathbf{r}_{ij}, \qquad 10^{-50} < \frac{r_{ij}}{r_{eq}} < \epsilon \tag{3.42}$$

and it is now applied over each dimension of the parametric space. The distance r_{eq} is that at which the velocity caused by the acceleration equals the mean velocity

---

[16]Today, the DANCe team is working on a GPU version of the code which is expected to compute the same amount of calculations in a couple of days.

caused by the common PSO. ϵ is a free parameter set, heuristically, to 10^{-10}.

3.6.2   Improvements of emcee

The modification I introduced in *emcee*, although very simple, improved the acceptance fraction and mixing of the particles. To allow the parallelisation, Foreman-Mackey et al. (2013) divide the ensemble of particles in two ensembles. In the original version, the particles in one ensemble use one and the same particle in complementary ensemble to compute their positions according to Eq. 3.39. In the modified versions, particles from one ensemble update their positions using a particle from the complementary ensemble, however, this particle is chosen randomly at each iteration.

Figure 3.5 compares the mixing and acceptance fractions of the original and modified versions of *emcee*. In a private communication with David Foreman, he mentioned that a similar modification was already introduced in a beta version of *emcee*.

Figure 3.5: Comparison between the log posterior evaluations of the original emcee version (left) and the modified one (right).

3.6.3   GMM for the field population

As mentioned earlier in this Chapter, the field population is modelled by means of two independent photometric and proper motion distributions. The MLE of the parameters of these distributions were found using the EM algorithm. The conventional EM algorithm for GMM (Dempster et al. 1977) for a mixture of M gaussians goes as follows. Given a set of parameters $\theta = \{w_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}_{i=1}^{M}$, were $w_i, \boldsymbol{\mu}_i$, and $\boldsymbol{\Sigma}_i$ are the fraction, mean and covariance matrix of gaussian component i, the likelihood of the data is,

$$p(\{\mathbf{y}_n\}_{n=1}^{N}|\theta) = \prod_{n=1}^{N} \sum_{i=1}^{M} w_i \cdot \mathcal{N}(\mathbf{y}_n|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \tag{3.43}$$

To solve the problem, the EM algorithm requires a set of N variables, z, of M dimension are introduced. The variable $z_{n,i}$ represent the probability that observation y_n was drawn from gaussian i. Therefore,

$$1 = \sum_{i=1}^{M} z_{n,i} \tag{3.44}$$

These z latent variables are found as

$$z_{n,i} = \frac{w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}{\sum_{i=1}^{M} w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)} \tag{3.45}$$

The EM works, as its name indicates, by maximising the expected value of the likelihood.

$$E[p(\{\mathbf{y}_n\}_{n=1}^{N} | \theta)] = \prod_{n=1}^{N} \sum_{i=1}^{M} z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \tag{3.46}$$

The previous expectation is maximal when,

$$w_i = \frac{1}{N} \sum_{n=1}^{N} z_{n,i}, \tag{3.47}$$

$$\boldsymbol{\mu}_i = \frac{1}{\sum_{n=1}^{N} z_{n,i}} \sum_{n=1}^{N} z_{n,i} \cdot \mathbf{y}_n \tag{3.48}$$

$$\boldsymbol{\Sigma}_i = \frac{1}{\sum_{n=1}^{N} z_{n,i}} \sum_{n=1}^{N} z_{n,i} \cdot (\mathbf{y}_n - \boldsymbol{\mu}_i) \times (\mathbf{y}_n - \boldsymbol{\mu}_i)^T \tag{3.49}$$

The modified version of the GMM, which includes a uniform distribution, is now a particular case of the GMM. This can be viewed as a gaussian distribution with fixed parameters and a constant probability c given by the uniform distribution. The new expectation is then,

$$E[p(\{\mathbf{y}_n\}_{n=1}^{N} | \theta)] = \prod_{n=1}^{N} \left[z_{n,0} \cdot w_0 \cdot c + \sum_{i=1}^{M} z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n | \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right]. \tag{3.50}$$

The maximisation step remains identical except for the indices. There are now $M + 1$ fractions w_i, with $i = 0, 1, ..., M$, and the means and covariances run from $i = 1, ..., M$.

Regarding the photometric GMM of the field. Since photometry has missing values, we use the EM algorithm of McMichael (1996). This algorithm was developed to obtain the MLE of data with missing values. A more recent version has been developed by Lin et al. (2006), however, this assumes that the missing data is randomly distributed (missing at random). As mentioned earlier, the missing values in the photometry are not randomly distributed.

In McMichael (1996) algorithm, there is a set of N gain matrices, one for each datum. Each M_n matrix is an identity matrix in which the rows of the corresponding missing value have been deleted. Thus, the expected value of the likelihood is now,

$$E[p(\{\mathbf{y}_n\}_{n=1}^N|\theta)] = \prod_{n=1}^N \sum_{i=1}^M z_{n,i} \cdot w_i \cdot \mathcal{N}(\mathbf{y}_n|M_n\boldsymbol{\mu}_i, M_n\boldsymbol{\Sigma}_i M_n). \tag{3.51}$$

The maximisation step is now,

$$w_i = \frac{1}{N} \sum_{n=1}^N z_{n,i}, \tag{3.52}$$

$$\boldsymbol{\mu}_i = \frac{\sum_{n=1}^N z_{n,i} \cdot H_n\mathbf{y}_n}{\sum_{n=1}^N z_{n,i} H_n M_n} \tag{3.53}$$

with

$$H_i = M_i^T (M\Sigma_i M^T)^{-1} \tag{3.54}$$

The maximisation step has no analytical solution for the covariance matrix, therefore, a modified steepest descent is used

$$\Sigma_i \leftarrow \Sigma_i + \frac{\rho}{2} \cdot \Sigma_i \Delta_i \Sigma_i, \tag{3.55}$$

with and Δ_i given by

$$\Delta_i = \frac{1}{\sum_{n=1}^N z_{n,i}} \cdot \sum_{n=1}^N z_{n,i} \cdot \left[H_n(\mathbf{y}_n - M_n\boldsymbol{\mu}_i) \times (\mathbf{y}_n - M_n\boldsymbol{\mu}_i)^T H_n^T - H_n I \right] \tag{3.56}$$

The algorithm preserves the monotonic convergence of the conventional EM,

and returns positive definite matrices provided that $\rho < 2$. For details and validation see McMichael (1996).

3.6.4 Hybrid-HPC implementations

As I outlined at the beginning of this section, the parallel computing approach was an unavoidable step. Once the code was ported to python and *CosmoHammer*, I modified this last one to better fit our needs. The modifications were mainly on the management of input and output files and the python *multiprocessing* package for the computing of our likelihood. Also, I striped some of its original functions to reduce memory usage and implemented some others, like the use of initial positions for the particles.

In our Hybrid-HPC (HHPC) approach, the particles of *emcee* are distributed on the nodes of the computing cluster. Then, each core in the node computes the likelihood of one fraction of the objects in the data set. For the cluster at the Centre of Astrobiology [17](CAB) I used a configuration of 6 nodes each with 12 cores. For the cluster at the University of Cádiz, I used a configuration of 5 nodes each with 16 cores. Particles were distributed among the six nodes using MPI while each core computes the likelihood of one 12th of the total number of objects in the data set.

However, the HHPC approach was not the best solution at the CIMENT clusters of the University of Grenoble Alpes. At this cluster (Froggy) the HHPC continuously render errors of communication. For this reason, I implemented a only MPI version of the code. In this version, the multithreading approach is left aside. Instead, the totality of the available cores is fully dedicated to the computing of the likelihood. Each of the n cores computes the likelihood of one nth of the objects in the data set. Once the likelihood of all particles has been computed, the master evaluates the new positions of the particles.

I finish this chapter with a brief description of the difficulties faced on the development and testing of the BHM code. As the code evolved in complexity, my computational skills were compelled to evolve as well. I started by learning R and solving some toy problems on it. Later, when the dimensionality of the

---

[17]Villanueva de la cañada, Madrid,Spain.

posterior increased I learned *Stan*. When the data set increased in size, we faced the parallelisation, so I learned Python and MPI. Once these versions were operable, I was forced to deal with libraries, from the common, numpy, scipy and numba, to the linking of modules and libraries. Finally, when dealing with several clusters I learned the queue languages Condor, slurm and OAR. Currently, the DANCe team has developed a GPU implementation of the BHM code.

Chapter 4

Results

In this Chapter I present the results of applying the methodology explained in Chapter 3 to the Pleiades DANCe data set (Sect. 2.7). However, to characterise the methodology and estimate its precision and accuracy, I first apply it over synthetic data. Using this data, I am able to analyse the performance of the methodology when it is considered only as a classifier. Later, in this Chapter I give the results of our methodology main objective: the statistical characterisation of the cluster population. In the following Sections I give the details of the spatial, velocity, luminosity and mass distributions. Then, I describe the physical scenario of the evolution of the mass distribution by comparing the Pleiades mass distribution with other younger and older clusters. Finally, I end this Chapter with a description of how the Bayesian methodology allowed us to update our previous knowledge of the Pleiades cluster.

4.1   Performance of the classifier

As mentioned earlier, the main objective of the methodology of the BHM is the statistical characterisation of the NYOC populations. As a by product, it also obtains individual membership probability distributions. Using these last ones, we are able to directly classify objects into cluster and field members, providing that an objective probability threshold has been established. The aim of this section is to explain the procedure to find the objective threshold by means of synthetic data. As any other measured property, this classification has an uncertainty, thus,

the purpose of this section is also to quantify this uncertainty.

That said, in order to properly characterise our classifier, I test it over synthetic data sets that resemble the real data. An ideal test to our classifier will be to apply it over well known dataset in which tags of cluster and field members were already present. However, if we may have access to these tags, a classifier may not be needed. The Pleiades cluster being *the* most studied cluster in history, it is the NYOC with most of these tags. This is the reason for which we decided to benchmark our methodology on it. Nevertheless, the problem of the synthetic data remains since the low-mass end of the cluster still is *terra ignota*. To over come this issue, we decided to create synthetic data sets under the assumption that our cluster and field models resemble the real data. We are aware that these models are far from perfect, but so far it is the best we can do.

The assumption that our methodology correctly models the real data, although enable us to measure the uncertainty of our classifiers, does not give any indication about possible biases in our model. To explore this possibility, we later compare our real data results with those found in the literature. I present this comparison at the end of this section.

Hence, we fit our models, field and cluster, to the real Pleiades DR2 data set consisting of the 10^4 high membership probability candidates according to Bouy et al. (2015b). Later, using the Maximum A Posteriori (MAP) estimates, we created synthetic stars (five samples of 10^4 objects each). To further test the reliability of our classifier, we compare the results it render when applied over data sets with and without missing values. This comparison allows us to quantify the impact that missing values have on our results.

One further consideration. The synthetic analysis requires at least three runs: one on the real data to obtain the MAP, and two on the synthetic one: with and without missing values. However, as explained in Chapter 3, our methodology is computationally expensive. Therefore, to maintain the computing time within reasonable limits (couple of weeks), we decided to restrict our synthetic data set to only the 10^4 objects with higher membership probability according to Bouy et al. (2015b). I elaborate on the consequences of this decision.

These 10^4 objects are "closer" to the cluster in the sense of membership probability than the remaining 9×10^4 objects. Therefore, the field probability density

rendered by this sample, compared to that of the larger 10^5 sample, has the following properties: it is more concentrated and has larger values near the cluster region. Since the 10^5 sample is largely dominated by the field population, its density peaks far form the cluster proper motion and photometric sequences. Densities are normalised, thus more of the density mass of this large sample is far from the cluster region.

Given the previous considerations, we assume that results obtained on the smaller 10^4 sample are more contaminated, and have lower recovery rates than those obtained on the larger and more distant 10^5 sample. On the one hand, the higher contamination rate results from the larger concentration of the field density around the cluster region. On the other hand, the lower recovery rates arises from the higher values of the field probability density. In simple words, when we define the field in a more restricted region around the cluster, both populations become more entangled, thus they become more difficult to separate. Therefore, we assume that the results obtained on the smaller 10^4 sample represent upper and lower limits to the contamination and recovery rates of the larger 10^5 sample, respectively.

Briefly, to create the synthetic data set, the procedure is the following. First, using the methodology of the previous Chapter, I obtain a sample of the posterior distribution of the parameters given the 10^4 real data set. Then, I chose the particle with highest posterior probability as the MAP estimate of the posterior distribution. Using this particle positions I generate five synthetic data sets of 10^4 objects each. Then, I tag these objects according to their parent population: cluster or field. Then, using the synthetically observed values I estimate their uncertainties and missing value patterns (more details below). Finally, I run the model over these five synthetic samples and compare the measured tags with the true ones as function of the probability threshold.

Also, I run the methodology over the synthetic data set without missing values and compare this results with those found on the same data set with missing values. This test, as mentioned before, enable us to quantify the impact of missing values over individual membership probabilities.

As explained in Sect. 2.7, our data set has a high fraction of missing values. Only $\sim 1\%$ has completely observed entries. Furthermore, the missing pattern is

not random and depends on the magnitudes and colours of the objects. Therefore, to better reproduce this pattern, for each synthetic datum, we use missing value template of one of its closer neighbours in the real data; closer in the euclidean sense. Using the missing value template of the nearest neighbour from the real data set results in a biased sample in which objects with complete (non missing) values are underestimated. This is the inevitable consequence of the fact that euclidean distances measured in subspaces resulting from the missing values, are smaller or, at most, equal to those measured in the non-missing value spaces.

Missing values are assigned as follows. Since by definition of our data set there are no missing values in our proper motion data, missing values were assigned only to photometry. We chose the closer neighbours from the available CMDs: $\{K_s, J - K_s\}, \{J, J - H\}, \{K_s, H - K_s\}, \{J, Y - J\}, \{K_s, i - K_s\}$. These CMD are formed with the bands and colours with fewer missing values, in decreasing order. The missing value pattern for individual objects was chosen as follows. First, for each CMD subspace we find the fraction, f_r, of objects from real data without missing values, we call it $C_{or,i}$. Then, we take a random sample from the synthetic data whose fraction, f_s corresponds to f_r. For objects in this sample we assign the missing value pattern of the nearest neighbour from sample $C_{or,i}$. We repeat the procedure for all CMDs. In this way, the synthetic data has fractions of missing and non-missing values similar to those of the real data.

Uncertainties are assigned as follows. We set the proper motions uncertainties to those of the nearest neighbour in the real data. In photometry, however, this scheme renders uncertainties that are biased towards the less precise measurements. This is a consequence of the missing values. Again, the euclidean metric results in the preferential choosing of objects with missing values. These missing values occur mostly at the faint end, where uncertainties are larger. Therefore, the uncertainties are biased towards larger values. To avoid this issue, we fit polynomials (8th degree) to the uncertainties as a function of the magnitudes. Then, we use these polynomials to give uncertainties to the synthetic photometric data.

The performance of our classifier was measured by counting the true positives (TP, cluster members correctly classified), true negatives (TN, field members correctly classified), false positives (FP, field members classified as cluster members) and false negatives (FN, cluster members classified as field members) recoveries as

a function of the probability threshold. With them we calculate the true positive rate, contamination rate, accuracy and precision, which are defined as follows. In order to classify the objects as cluster of field members we summarise their membership probability distribution using the mode. If the mode is greater than the current probability threshold, then the object is classified as cluster member, if not as field.

The true positive rate (TPR) is the ratio of true positives over the sum of true positives plus false negatives. The contamination rate (CR) is the ratio of false positives over the sum of false positives plus true positives. The precision or positive predictive value (PPV) is the ratio of true positives over the sum of true positives plus false positives. Finally, the accuracy (ACC) is the ratio of the sum of true positives plus true negative over the sum of true and false positives and negatives. These are,

$$TPR = \frac{TP}{TP + FN}$$
$$CR = \frac{FP}{FP + TP}$$
$$PPV = \frac{TP}{TP + FP}$$
$$ACC = \frac{TP + TN}{TN + FN + TP + FP}$$

We use the results of the five synthetic data sets to quantify the uncertainties of the previous quantities.

In Fig. 4.1 I show the TPR and CR, as a function of probability threshold, computed for the cases in which missing values where both present and absent. The missing value case was computed for the five synthetic samples, thus, the lines and the shaded grey regions depict the mean and maximal deviations, of the results found on the five synthetic samples. As can be seen in this Figure, the missing values have a negative impact in our classification. They cause a diminishing of the TPR and an increase in the CR. This negative impact is expected since the observables we are using are highly discriminant in the classification process. Since cluster and field are highly entangled in our 10^4 synthetic samples, when a highly discriminant observable is missing the classification process could be biased or more

uncertain.

In spite of the negative impact of missing values, our methodology delivers low ($\lesssim 8\%$) contamination rates and high recovery rates. In Figure 4.1 we also show, for the sake of comparison, the CR and TPR of Sarro et al. (2014) (reported in their Table 4).

From the previous comparison we observe that the TPR delivered by of our methodology: i) when measured on data without missing values, is similar to that of Sarro et al. (2014) methodology. This is expected since the results of those authors are based on a model constructed only with completely observed objects (i.e. non-missing values). ii) when measured on data with missing values, is $\approx 4\%$ lower than that of Sarro et al. (2014).

This last figure is a small price to pay compared to the dramatical increase in the number of sources used to construct our model (both field and cluster). As mentioned in Section 2.7, in our 10^5 objects data set, only $\sim 1\%$ of objects have completely observed entries. Roughly, this is the fraction of sources used by Sarro et al. (2014) and Bouy et al. (2015b) to construct their models.

On the other hand, the CR of our methodology, above $p = 0.8$ and for both missing-values and non-missing values cases, outperform, by $x\%$, the CR reported of Sarro et al. (2014) methodology, see below.

In simple words, what can be observed from the previous analysis is that our methodology, compared to that of Sarro et al. (2014), renders less contaminated (CR) results ($x\%$) at the price of a smaller recovery rate (TPR) 4%

Nonetheless, we stress the fact that the previous comparison is not straight forward. The following reasons must be born in mind. First, Sarro et al. (2014) infer their cluster model using only non-missing-value objects, later they apply that model to objects with and without missing values. Second, their synthetic data set and ours are essentially different. They are constructed with different generative models, different number of elements, and different missing value patterns.

Now, I describe the procedure to set an objective probability threshold. This probability threshold, although not needed to obtain the distribution of the cluster population, is needed however to objectively classify an object. Since real data contain missing values, there is no need to establish this threshold for the non-missing values case. There are different approaches to establish this probability

threshold, however I only use the approach of maximum accuracy (ACC).

Figure 4.1 shows the ACC and the PPV of our classifier when applied on synthetic data with missing values. The lines and the grey regions depict, respectively, the mean and the maximum deviations of the five synthetic data set results. We use these last ones as the uncertainty of our estimates. The mean of the highest accuracy, ACC$=96.5 \pm 0.1\%$, happens at probability threshold $p_t = 0.84$. At this threshold the CR is $4.3 \pm 0.2\%$, the TPR is $90.0 \pm 0.05\%$, the ACC is , and the PPV is $95.6 \pm 0.2\%$.

Figure 4.1: Left: The TPR (solid line) and CR (dashed line) of our methodology when applied on synthetic data sets with and without missing values (red and blue lines, respectively). In black dots we show the TPR and CR reported by Sarro et al. (2014) for their non-missing values model. Right: Accuracy and precision as a function of probability threshold for our classifier when applied on synthetic data with missing values. The higher accuracy is obtained at $p_t = 0.84$ (red dot). In both panels, the grey areas show the maximum deviations from the mean of the results of the five missing-values synthetic data sets.

We investigate further on the impact of missing values, particularly to analyse any possible biases introduced by them. It was done by comparing the mode of the individual membership probability distributions found after fitting the model to data sets with and without missing values. The missing values data set is the result of adding a mask of missing values (as previously described) to the completely observed data set. It means that this data set has identical observed values, except for those masked as missing ones.

In Fig. 4.2 we show the mode of the membership probabilities of objects with missing values (vertical axis) against those of non-missing values (horizontal axis). As can be seen in this Fig., the missing values impact our results by spreading the membership probabilities. Ideally, we would like to recover membership probabilities following the line of slope one, as in the case of completely observed values (red squares). The most striking deviations come from objects lacking the CI (enclosed in black). Our methodology uses the *true* CI to prescribe the *true* photometry, and the observed CI to constrain the marginalisation integral of the *true* CI. Thus, it is expected that a missing CI will produce a probability spread. These missing CI

objects show two different behaviours. In one case, there are sources with membership probabilities $p_{complete} \approx 0$ which have overestimated probabilities in the incomplete case (vertical axis). In the other case, the sources in the combed area below the line of unit slope have underestimated probabilities in the incomplete case. While the first case contributes to the CR the second one diminishes the TPR. The first case reaches the maximum difference at $p \approx 0$ (difference between red and blue dashed lines in Fig. 4.1), thus its impact in our results is marginal. Furthermore, at our objective probability threshold $p_t = 0.84$, these objects have a small impact in our results, they represent only 1.8% of the contamination rate. The box region in Fig. 4.2 shows them. The second case, however, typify the unavoidable loss of members due to the missing values, 4% at $p_t = 0.84$, given our model and the observables it uses.

The bias introduced by missing values can be estimated using the root-mean-square (rms) of the difference between membership probabilities recovered with and without missing values. The total rms is 0.12. On the one hand, objects with completely observed values (red squares) in both data sets have a rms of only 0.02. On the other hand, objects with missing values have a rms of X. While the rms of objects lacking the CI have and rms of X. The previous effects show an overall agreement between results on data sets with and without the missing values, nonetheless, care must be taken when dealing individually with objects lacking the CI.

As mentioned before, our methodology aims at the statistical distributions of the cluster population. It works by ensuring that each object contributes to the cluster distributions proportionally to its cluster membership probability. In this sense our results are free of any possible bias introduced by hard cuts in the membership probability. Nevertheless, contamination is still present and must be quantified. To quantify it, I compute the expected value of the CR. It is $\langle CR \rangle = 5.8 \pm 0.2\%$. In this expected value, each CR contributes proportionally to the probability threshold at which it is measured.

In machine learning and data mining is sometimes useful to analyse the performance of a binary classifier by the receiver operating characteristic curve, the ROC curve. This plot is a visual diagnostic of the ability of a classifier to to its job. The ROC curve plots the TPR as a function of the FPR for all possible values of the

Figure 4.2: Comparison between the cluster membership probabilities recovered from the synthetic data with missing values (Incomplete) and without them (Complete). The colour and shape indicate the amount of missing values. The symbols enclosed in black indicate a missing CI. The top left box contains objects considered as contaminants due to missing values, at the probability threshold $p_t = 0.84$.

probability threshold. A perfect classifier would be that in which the TPR=1 and the FPR=0 for some probability threshold. On the other hand, a random classifier would be that on which the TPR=FPR at all probability thresholds. Such classifier has a line of slope one as its ROC curve. Furthermore, the quantitative diagnostic for a binary classifier is the area under the ROC curve (AUC). As its name indicate, the AUC is the integral of the ROC curve. Thus a random classifier has a AUC of one half, while a perfect classifier as AUC=1. In Fig. 4.3, I show the ROC curve for our classifier when applied over synthetic data with missing values. It is the ROC of one of the five synthetic realisations described throughout this section. As can bee seen from this Fig. our classifier does an excellent job, with an AUC=0.99.

Figure 4.3: ROC curve of our classifier when applied over one of the synthetic data sets described earlier. As can be seen the AUC=0.99 diagnose it as an excellent classifier.

4.1.1 Comparison with the literature

In Chap. 2 I mentioned that the most important works on the Pleiades members are those of Stauffer et al. (2007); Lodieu et al. (2012); Sarro et al. (2014); Bouy et al. (2015b). In this section I will compare the list of candidate I members found with the most recent study of the Pleiades members, that of Bouy et al. (2015b). Later, for the sake of completeness, I will compare my list of candidate members with the compilation made by Stauffer et al. (2007).

4.1.1.1 Candidate members from Bouy et al. (2015b)

The methodology of this work, although essentially different from that of Bouy et al. (2015b), the fact that both share the data set and use the same observables, allows a direct comparison between them. When using their objective probability thresholds, as shown by Fig. 4.4, both methodologies agree on the outstanding 99.6% of the classified objects. Concerning the candidate members, they also agree on $\approx 90\%$ of them. Nevertheless, the discrepancies are worthy of discussion.

The rejected candidates of Bouy et al. (2015b) (lower right box of Fig. 4.4) amount to 12% of their total number of candidate members. This value is 4.7% higher than the contamination rate reported by Sarro et al. (2014): $7.3 \pm 1.4\%$. I will not reject a priori these objects as candidate members because, as mentioned in Sect. 4.1, our methodology recovers 4% less members when compared to that of Sarro et al. (2014)(see Fig. 4.1). The majority, 85%, of these objects have missing values, which indicates that these objects require future follow up. Even the 37 completely observed and rejected candidates can not be discarded as true members due to the fact that our methodology losses 6% of the true cluster members (see Fig. 4.1).

On the other hand, our new candidates (upper left box of Fig. 4.4) amount to 10% of Bouy et al. (2015b) candidates. This figure is higher than the $4.3 \pm 0.2\%$ CR reported on the previous Section. Thus, it indicates that up to $\sim 6\%$ of these objects can be true cluster members. From these objects, one half have completely observed values, thus 3% of our new candidates are completely observed objects. This last figure agrees perfectly well with the 3% of non recovered members reported by Sarro et al. (2014).

In Figs. 4.5 and 4.6 I show the proper motions and K_s vs $i-K_s$ CMD projection spaces of our new candidates and the rejected ones of Bouy et al. (2015b). In the following I elaborate on their properties.

The new candidate members have proper motions uncertainties whose median, $\overline{\mu_{\alpha,\delta}} = \{1.33, 1.33\}\ mas/yr$, is two times larger than those of the candidate members in common with Bouy et al. (2015b), median $\overline{\mu_{\alpha,\delta}} = \{0.65, 0.65\}\ mas/yr$. Also, as shown by Fig. 4.5, the majority of the new candidate members, 148, have probabilities lower than 0.95, are located in a halo around the locus of the cluster proper motions, and on top of the cluster sequence in the K_s vs $i - K_s$ CMD. On

the contrary, the new candidates with probabilities higher than 0.95, which are 39, lay in the centre of the cluster proper motions and fall above the cluster sequence in the K_s vs $i - K_s$ CMD. Thus, I hypothesise that: i) Objects whose photometry is compatible with the cluster sequence but are in the proper motions halo, have higher membership probabilities in our methodology due to the increased flexibility of the cluster proper motions model: it now has four gaussians instead of the two of Bouy et al. (2015b). And ii) objects near the centre of the cluster proper motions but located above the cluster photometric sequence sequence, are multiple systems (probably triple systems which can amount to 4% of the population Duquennoy & Mayor 1991) with an increased membership probability due to our more flexible photometric model of the cluster and equal-mass binaries sequences.

The rejected candidates of Bouy et al. (2015b), as it is shown in Figs. 4.6 and 4.7, have proper motions uncertainties with median $\overline{\mu_{\alpha,\delta}} = \{3.15, 3.19\}$ mas/yr. This value is more than four times larger than that of the candidates in common with our members. Also, these objects are distributed along the cluster sequence. Among these objects, those with a relatively high membership probability occur mostly at the middle of the cluster sequence (green squares of Fig. 4.7) while those with lower membership probabilities occur at the bright and faint ends (blue and red triangles of Fig. 4.7, respectively). These last regions coincide with those where the missing values happen the most. We stress the fact that Sarro et al. (2014) and later Bouy et al. (2015b) construct their models using only completely observed objects (i.e. those without missing values). For their field models both authors use a sample of $\approx 20,000$ objects. Proceeding in that way, as explained in Sect. 3.3.1, underestimates the field density, particularly in the regions where the missing values are more frequent. Underestimating the field likelihood increases in the cluster field likelihood ratio, therefore it increases also the cluster membership probabilities. Furthermore, the proper motions uncertainties of objects at the bright middle and faint ends, have medians of $\overline{\mu_{\alpha,\delta}} = \{4.0, 4.2\}$ mas/yr, $\overline{\mu_{\alpha,\delta}} = \{2.4, 2.4\}$ mas/yr and $\overline{\mu_{\alpha,\delta}} = \{3.4, 3.4\}$ mas/yr, respectively. These figures are approximately 6, 4 and 5 times larger, respectively, than those of the candidates in common. These large uncertainties produce a proportional spread of the likelihood distribution thus further reducing the membership probability.

I consider that both the large proper motion uncertainties and field likelihoods

are responsible for the diminished membership probabilities of Bouy et al. (2015b) rejected candidates. However, these rejected candidate members cannot be discarded as potential members. Indeed, at the probability threshold of maximum accuracy, $p_t = 0.84$, the TPR is just $90.0 \pm 0.05\%$. It means that there are still 10% of true members within the rejected candidates. To rule out the possibility that these objects are indeed members we need lower proper motion uncertainties and fewer missing values. Future steps will be taken to try to solve this issue.

Summarising, the discrepancies between our individual membership probabilities and those reported by Bouy et al. (2015b) arise from subtle but important differences. The first of them is the more formal treatment of missing values in our methodology and its inclusion in the field model. Taking into account the missing values has two main consequences. The first of them is that, the new photometric model of the field diminished membership probabilities, particularly in the regions where missing values happen the most. Second, the use of missing values in the construction of the cluster model allow us to include the information of good candidate members that were otherwise discarded a priori. The second difference is the higher flexibility of our cluster model, it allows us to increase the membership probability of the previously discarded candidates. Furthermore, as shown by the red squares in the upper left corner of Fig. 4.4, the higher flexibility of our cluster model allow us to include as new candidate members previously rejected objects with complete (non-missing) values.

Figure 4.4: Recovered membership probabilities compared to those of Bouy et al. (2015b). Lines show the 0.75 and $p_t = 0.84$ probability thresholds used in both works. The numbers indicate the new candidate members (top left), rejected candidate members (bottom right), and common candidate members (top right).

Figure 4.5: Proper motion (left) and K_s vs. $i - K_s$ CMD (right) showing the new candidate members found in this work. Captions as in Fig. ??.

Figure 4.6: Proper motion (left) and K_s vs. $i - K_s$ CMD (right) showing the rejected candidate members of Bouy et al. (2015b). Captions as in Fig. **??**.

Figure 4.7: Proper motion (left) and K_s vs. $i - K_s$ CMD (right) showing the rejected candidate members of Bouy et al. (2015b). The colours and shapes are a proxy for their K_s magnitude.

4.1.1.2 Candidate members from Stauffer et al. (2007)

.

Stauffer et al. (2007) published two list of candidate members. The first one contains 1417 objects compiled from the literature (their Table 2). These objects were classified as candidate members by several authors along modern astronomy. As Stauffer et al. (2007) mention, this list is inhomogeneous, incomplete and certainly includes non members. Their second list contains 55 candidate members (their Table 5) found using infrared photometry, and proper motions.

Cross matching these two lists with the candidate members ofBouy et al. (2015b), shows that only 1288 of the 1417 objects on Stauffer et al. (2007) Table 2 were classified by Bouy et al. (2015b) as candidate members. From these, 1132 objects come from the DANce survey and the rest from Tycho+DANCe, see Appendix B of Bouy et al. (2015b). Also, only 34 of the 55 new candidate members of Stauffer et al. (2007) were classified by Bouy et al. (2015b) as candidate members. In Figures 4.8 - 4.9 I show the candidate members of Stauffer et al. (2007), and Bouy et al. (2015b). As mentioned by Stauffer et al. (2007), many of the candidate members in their Table 2 are most probably non members. These objects lay far from the locus of the rest of the candidates. Similarly, the majority of the rejected candidates of Stauffer et al. (2007), 15 of those in their Table 5, lay far from the cluster proper motion and photometric loci.

Concerning our list of candidate members, after cross matching the lists, we recover 1139 of the candidate members from Table 2 of Stauffer et al. (2007). Seven more than Bouy et al. (2015b). Also, we recover the same 34 candidates that Bouy
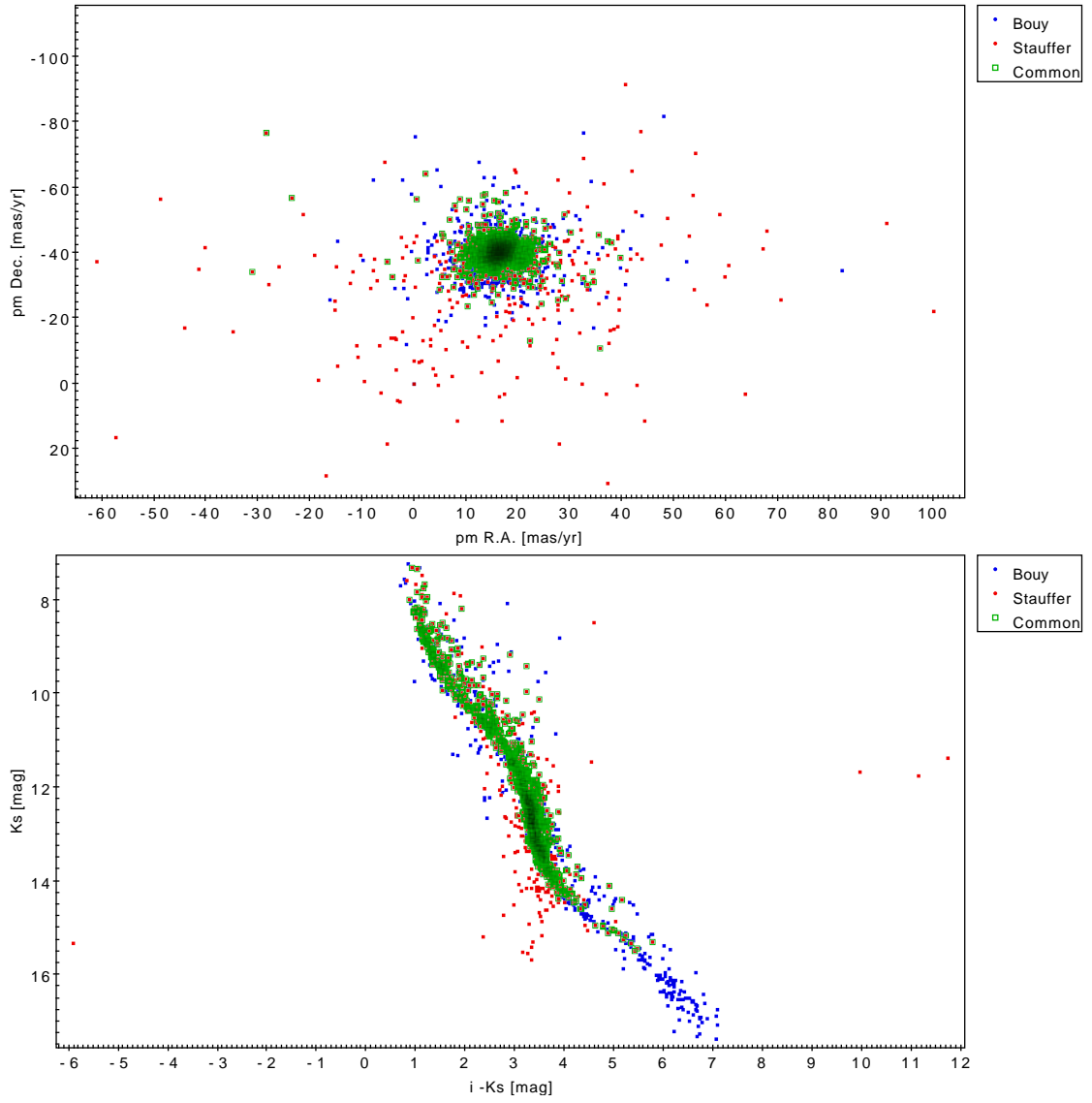
Figure 4.8: Comparison between proper motions (top) and K vs $i - K$ CMD of the candidate members compiled by Stauffer et al. (2007) (red) and those of Bouy et al. (2015b)(blue), the common candidates are shown within green squares.
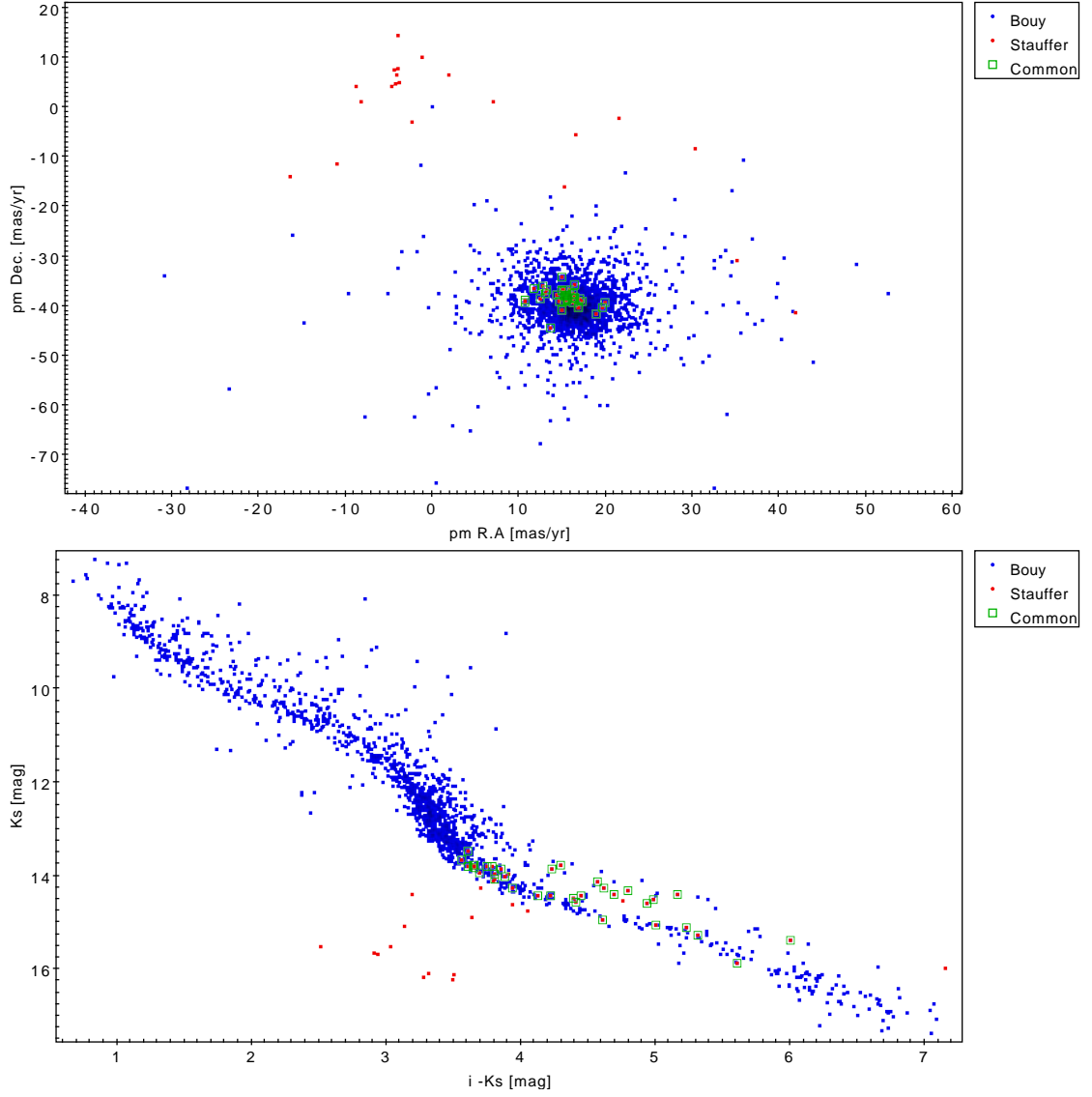
Figure 4.9: Comparison between proper motions (top) and K vs $i - K$ CMD of the new candidate members of Stauffer et al. (2007) (red) and those of Bouy et al. (2015b)(blue), the common candidates are shown within green squares.

et al. (2015b) recovered from the new candidates of Stauffer et al. (2007) (from their Table5).

I have already mentioned that the Table 2 of Stauffer et al. (2007) is an exhaustive compilation of Pleiades members. This list contains objects that authors from the literature classified as Pleiades members even when the membership probability was as low as 0.1 (Stauffer et al. 2007). Therefore, I will not analyse in detail the ~ 300 rejected objects. It suffices to mention that we recover an slightly larger (seven) number of candidates than Bouy et al. (2015b).

On the other hand, the list of new candidate members of Stauffer et al. (2007) deserves further attention and a more detailed comparison. From the 21 rejected objects, 14 of them lay below the cluster photometric sequence and far from the proper motion locus. The remaining seven have membership probabilities near our probability threshold. XXXXX CHECK THIS XXXX

4.2 The statistical distributions of the Pleiades cluster.

Once the objective probability threshold has been established and the results of the classification analysed and compared with the literature. Now I present the results of the statistical distributions that describe the cluster population. These distributions result directly or indirectly from the posterior distribution of the parameters in our model. Here I summarise these parametric posterior distributions, together with some of their correlations. Also, I present and analysis on the way these posterior distribution help us to update our prior knowledge.

In Table **??** I summarise the posterior distribution of the parameters in our model. I use as statistic and uncertainty the mode and the 16th and 84th percentiles, respectively. The parameter names correspond to those given in Sect. **??**.

4.2.1 Updating the prior knowledge

As mentioned by Gelman (2006), the posterior distribution must be inspected to update our previous knowledge. To inspect these posterior distribution, I use the

bare values and the correlations among them. The values indicate, for example, that the number of GMM modelling the proper motions of the single stars is overestimated. The fraction and variance of the last gaussian went both to near zero values. A better model would be that in which no computing power will be lost in inferring negligible parameters.

Another simple inspection tells us that our prior for the cluster field fraction is narrow and its mode lay far from that of the posterior distribution. Although this is not wrong since the prior clearly allow this value, this prior must be made wider. Clearly our prior knowledge underestimated the expected number of cluster members.

Another useful tool to inspect these posterior distributions is the correlation they show with themselves. In Figure 4.10 I show the correlation matrix of all the parameters in our model. This Fig. shows elements that could be considered to improve our model. For example, the large correlation shown by the high order coefficients of the spline series may indicate that we can save some of these parameters.

Figure 4.10: Correlation matrix of parameters in our model. The colour code indicates the value of the correlation coefficient.

In the following Sections, I use the distributions of the parameters in our model to derive the statistical distributions that describe the cluster population. Some of these distributions have as parameters those inferred in the model, however, other, like the mass distribution require more elaborated derivations.

Discuss how to deal with future clusters in which we will not have Bouy2015 for the priors.

IS BIC a good approximation, what can be done in the future?

4.3 Velocity distribution

I must work on it. Here I must transform the bivariate proper motions distributions into a univariate velocity distributions using the distance and assuming

spherical symmetry. This distribution will give us some hints to improve the proper motions model. It will be interesting to fit a model, maxwellian for example and see how well it fits. If more models are available we can try to find the best one in terms of Bayesian model selection.

4.4 Spatial distribution

To include

4.5 Luminosity distribution

This Section describes the process by which the apparent J, H, and K_s magnitude distributions are obtained and then transformed into the absolute magnitude distributions. Later, I present these distributions and compare them with those found by Bouy et al. (2015b). Also I will compare them with the ones resulted from only the high membership probability objects.

4.5.1 Derivation of the magnitude distributions

To derive the J, H, K_s magnitude distributions I start with the colour index, CI, distribution. This last one, is described by an univariate GMM whose parameters are inferred in the model. Since we also model the EMB, their fraction and photometric sequence are taken into account.

In the following, I exemplify the process of derivation on the K_s band. Similar transformations apply to the rest of the bands. To obtain the distribution of K_s for the cluster objects, I take the colour index CI as a nuisance parameter, later I marginalise it. Thus,

$$p(K_s|\boldsymbol{\theta}_c) = \int p(K_s, CI|\boldsymbol{\theta}_c) \cdot dCI = \int p(K_s|CI, \boldsymbol{\theta}_c) \cdot p(CI|\boldsymbol{\theta}_c) \cdot dCI.$$

The term $p(K_s|CI, \boldsymbol{\theta}_c)$ corresponds to the GMM modelling the distribution of CI (Eq. 3.28), while $p(K_s|CI, \boldsymbol{\theta}_c)$ is the probability of K_s given the CI, and the

cluster parameters $\boldsymbol{\theta}_c$. The EMB are included with an amplitude equal to their fraction, $(1 - \pi_{CB})$. Thus,

$$
\begin{aligned}
p(K_s|\boldsymbol{\theta}_c) = \int & [\pi_{CB} \cdot p_{Cs}(K_s|CI, \boldsymbol{\theta}_c) + (1 - \pi_{CB}) \cdot p_{Bs}(K_s|CI, \boldsymbol{\theta}_c)] \\
& \cdot p_{CI}(CI|\boldsymbol{\theta}_c) \cdot dCI. \\
= & \pi_{CB} \int p_{Cs}(K_s|CI, \boldsymbol{\theta}_c) \cdot p_{CI}(CI|\boldsymbol{\theta}_c) dCI \\
& + (1 - \pi_{CB}) \int p_{Bs}(K_s|CI, \boldsymbol{\theta}_c) \cdot p_{CI}(CI|\boldsymbol{\theta}_c) \cdot dCI.
\end{aligned}
$$

$$(4.1)$$

In this equation, Cs and Bs stand for cluster and EMB sequences, respectively. The terms inside the integrals correspond to Equations 3.27 and 3.28. However, since here I want to derive only the distribution of K_s, I marginalise the rest of the bands. Also, the integration limits must change to those of the truncated colour distribution ($CI_{min} = 0.8, CI_{max} = 8$). Hence,

$$
\begin{aligned}
p(K_s|\boldsymbol{\theta}_c) = \pi_{CB} \int_{CI_{min}}^{CI_{max}} & \left[\left[\sum_{i=1}^{5} \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}) \right] \right. \\
& \left. \cdot \int_{\tilde{Y},\tilde{J},\tilde{H}} \mathcal{N}(\{CI, \tilde{Y}, \tilde{J}, \tilde{H}, K_s\}|\boldsymbol{S}(CI, \boldsymbol{\beta}), \Sigma_{clus}) \ d\tilde{Y} \ d\tilde{J} \ d\tilde{H} \right] \cdot dCI \\
+ (1 - \pi_{CB}) \int_{CI_{min}}^{CI_{max}} & \left[\left[\sum_{i=1}^{5} \pi_{CI,i} \cdot \mathcal{N}_t(CI|\mu_{CI,i}, \sigma_{CI,i}) \right] \right. \\
& \left. \cdot \int_{\tilde{Y},\tilde{J},\tilde{H}} \mathcal{N}(\{CI, \tilde{Y}, \tilde{J}, \tilde{H}, K_s\}|T_{Bs}(\boldsymbol{S}(CI, \boldsymbol{\beta})), \Sigma_{clus}) \ d\tilde{Y} \ d\tilde{J} \ d\tilde{H} \right] \cdot dCI.
\end{aligned}
$$

The derivations of the J and H magnitude distributions are similar to the procedure described for K_s. It is worth of notice that, the derivation process takes into account the EMB and the systems (binaries or multiples) which could have different mass ratios. Therefore, we call them the system magnitude distributions.

The previous magnitude distributions, together with the parallax and extinction of the cluster, allows us to obtain the system luminosity distributions. Here,

I assume that the parallax is normally distributed with mean, 7.44 mas, and standard deviation 0.42 mas (Galli et al. 2017). This parallax distribution is then convolved with the magnitude distributions to obtain the absolute magnitude distributions. Finally, I deredden them employing the canonical value of extinction: $A_v = 0.12$ (Guthrie 1987). This last values were transformed into the J, H, K_s extinctions using the extinction law of Cardelli et al. (1989).

Our methodology prescribes the *true* photometric quantities based on the *true* colour index CI. Therefore, the completeness limits of this CI dictate those of the photometric bands. The upper completeness limits that Bouy et al. (2015b) estimate for i and K_s are $i \approx 23$ mag and $K_s \approx 18$ mag (see their appendix A). As these authors mention, due to the heterogeneous origins of the DANCe DR2 survey, the completeness is not homogeneous over its entire area. To overcome this issue, they identified a region, the inner three degrees of the cluster, with homogeneous spatial and depth coverage. Then, they restricted their analysis to this region.

Restricting the sample means that good candidate members are not taken into account. Furthermore, if any dynamical process has been set on the cluster such that the mass distribution of its members is not uniform in the space, then restricting the sample could introduce a bias. One of such dynamical process is mass segregation, which, as suggested by Adams et al. (2001b) may have happen in the Pleiades.

Instead of restricting the sample, I assume that the UKIDSS survey, which is the most profound of the DANCe catalogue, provides the homogeneous spatial coverage at faint magnitudes, thus providing the lower limit to the completeness. Then, for the bright end of the survey, I quote conservative completeness limits. Figure 4.11 shows the K_s and i density for all sources in the Pleiades DANCe DR2. The upper completeness limits correspond to the point with maximum density, $i = 21.4$ mag, $K_s = 18.1$ mag. The density at bright magnitudes shows a sharp decline, probably due to saturation. Therefore, I choose $i = 13.2$ mag and $K_s = 11.0$ mag as the lower completeness limits.

The CI completeness interval is then defined as that of all the points, along the cluster sequence in the K_s vs. $i-K_s$ CMD, for which i and K_s are bounded by their upper and lower completeness limits, respectively. This results in a completeness

interval for $i - K_s$ of $2.7 < i - K_s < 5.6$ mag. With this completeness interval and the cluster sequence (the splines), we derive the completeness intervals for the J, H, K_s. Finally, these intervals were transformed to absolute magnitudes and deredden.

Figure 4.11: Density of all DANCe DR2 sources in K_s vs i magnitudes. Lines show our completeness limits, $13.2 < i < 21.4$ mag and $11 < K_s < 18.1$ mag. The grey area is considered incomplete.

.

In Fig. 4.12, I plot the J, H, K_s luminosity distributions together with their completeness limits, hereafter I call these distributions the continuous BHM. For the sake of comparison I also show the following luminosity distributions. The luminosity distributions of objects whose mode of membership probability is greater than our probability threshold $p_t = 0.84$, I call these distributions the discrete BHM. Also, I plot the luminosity distribution resulting from the candidate members of Bouy et al. (2015b), I call it discrete Bouy. Since the discrete luminosity distributions, both Bouy and BHM, relay on the individual object magnitudes, and many of these objects have missing values, I impute them using the nearest euclidean neighbour.

The difference between the continuous BHM and discrete BHM distributions comes essentially from the objects used to obtain them. The continuous one uses all objects proportionally to their cluster membership probability while the discrete BHM uses only the high probability candidate members ($p > p_t$). Since the discrete BHM is not a random sample of the continuous BHM, therefore their distributions does not need to be exactly alike. In addition, the missing values in the continuous BHM case were marginalised, while in the discrete BHM were imputed.

On the other hand, the differences between the discrete distributions, BHM and that of Bouy et al. (2015b), arise mainly at the bright and faint ends ($K_s \approx 4$ mag and $K_s \approx 11$ mag). We argue that the origin of these differences lay in our new candidate members and in the rejected ones of Bouy et al. (2015b), as it is discussed in Sect. 4.1.1.1.

Figure 4.12: Luminosity functions from J, H, K_s (orange). Also shown the regions of incompleteness and the luminosity functions computed from: the candidate members of Bouy et al. (2015b) (dot-dashed blue line), and our candidate members, ($p_{84\%} > p_t$, dashed black line).

.

4.6 Mass distribution

This Section starts with a brief description of the mass-luminosity relation, which is used to derive the present day system mass function (PDSMF) of the Pleiades together with the luminosity distribution of the previous section. Then I compare this PDSMF to the Initial Mass Functions (IMFs) of Chabrier (2005) and Thies & Kroupa (2007). Later, I conclude this section with an analysis of some simple models that can be fitted to our the PDSMF, I give the best model according to the Bayesian evidence.

4.6.1 The mass-luminosity relation

The mass-luminosity relation is the key to transform luminosities into masses. This relation relies entirely on the current models of stellar atmosphere and evolution. Among the different flavours of models, we choose the BT-Settl model of Allard et al. (2012). We based this decision on the fact that currently, this is the only model which covers our luminosity range at the age of the Pleiades. Since the DANCe survey uses as reference the J, H and K_s bands of the 2MASS survey, I choose the BT-Settle grid: CIFIST2011bc for the 2MASS AB photometric system. This grid, as it name indicates, returns values of the luminosity on a (non-uniform) grid on the mass. However, the transformation of luminosities into masses is proportional and thus very sensitive to the Jacobian of the transformation. To avoid the discontinuities in the derivatives produced by the grid, we decided to fit the grid using spline series and then obtain the derivative from these series. It is important to notice that we implicitly assume that the transformation from luminosities to masses does not have any associated uncertainty, thus it is uniquely determined. We can not do more since the models do not provide any uncertainty.

4.6.2 Present day mass function

The PDSMF is independently obtained in the J, H, K_s bands by transforming the luminosity functions into system mass functions using the mass-luminosity relations described in the previous section. Since the luminosity functions of Sect. 4.5 correspond to the luminosity of systems (single and binary stars altogether), therefore the derived mass function corresponds to the system mass function. We assume an age of 120 Myr for the Pleiades together with solar metallicity, these are the canonical values reported by and , respectively.

Figure 4.13 shows the logarithmic PDSMF (ξ_L) for the J, H, K_s bands normalised on the completeness limits obtained in Sect. 4.5. This figure also shows, the PDSMF proposed by Bouy et al. (2015b) and, the IMFs of Thies & Kroupa (2007) and Chabrier (2005). For this last IMF, I show the standard uncertainty as the value reported by Chabrier (2003). As shown in this Fig., our PDSMFs compares well, in the completeness interval, with the one proposed by Bouy et al. (2015b). However, there are discrepancies, particularly above $0.3 M_\odot (-0.5 < \log M/M_\odot)$. These may have its origin on the higher membership probability of our new candidate members. These new candidate members are preferentially M stars, whose masses are the range $0.075 - 0.6 M_\odot$ $(-1.12 < \log M/M_\odot < -0.22)$. Furthermore, these differences root also on the objects that Bouy et al. (2015b) did not include in their analysis: those lying outside the inner three degree region.

To obtain a model for our PDSMF I proceed as follows. First, I select three models: a log-normal distribution and two power-law distributions of the form $m^{-\alpha}$, with two and three power-law segments. Second, I took the mode distribution of our 100 sample distributions of the PDSMF in the K_s band and completeness interval and draw a 10^4 synthetic sample. Afterwards, using *PyMultiNest* (Buchner et al. 2014), and the synthetic sample I fitted the three models. Finally, in Table 4.1, I give the posterior distributions of each model parameters together with its evidence (see Sect. 3.1.2).

Judging by the evidences, the best model corresponds to the two segment power-law distribution. This model is over plot as solid black line in Fig. 4.13. This model agrees well with that found by Bouy et al. (2015b), except for the flat part in the low-mass range and the less steep slope in the high mass range. Nevertheless,

it is in clear discrepancy with the IMFs of Chabrier (2005), ($m_c = 0.25^{+0.021}_{-0.016}$ and $\sigma = 0.55^{+0.05}_{-0.01}$, the uncertainties are those reported by Chabrier 2003, for single objects) and of Thies & Kroupa (2007). The discrepancy between the IMFs and the PDSMFs (Bouy et al. 2015b, and ours) may have its origin on the not yet established uncertainties in the mass-luminosity relationship, on dynamical effects associated with age, or on both of them. In the next section I compare the PDSMF of the Pleiades with that of other younger and older clusters in order to analyse if there is substantial evidence to claim for dynamical effects associated with age.

Table 4.1: Parameters and evidence of models fitted to the PDSMF

Model	Parameters	Log Evidence
LogNormal	$m_c = 0.36 \pm 0.03$	
	$\sigma = 0.46 \pm 0.02$	18.1 ± 0.1
Two Segments	$\alpha_0 = -0.11 \pm 0.06 \quad m \in [0.04, 0.22 \pm 0.01]$	
	$\alpha_1 = 1.13 \pm 0.1 \quad m \in [0.22 \pm 0.01, 0.56]$	2222.7 ± 0.4
Three Segments	$\alpha_0 = -0.05 \pm 0.6 \quad m \in [0.04, 0.08 \pm 0.03]$	
	$\alpha_1 = -0.1 \pm 0.1 \quad m \in [0.08 \pm 0.03, 0.22 \pm 0.01]$	
	$\alpha_2 = 1.13 \pm 0.1 \quad m \in [0.22 \pm 0.01, 0.56]$	2221.2 ± 0.3

Before concluding this section, I use our PDSMF to give a lower limit to the mass of the cluster. The cluster members mean mass, in our entire mass range, is $0.26 \pm 0.006 M_\odot$. Thus, the product of this mean mass with the expected number of cluster members, 3116 ± 110, is an estimate of mass of the cluster: $807^{+38}_{-29} M_\odot$. The expected number of cluster members is the integral, over the whole range of membership probabilities, of number of objects at each membership probability. However, since we still lack the low and high mass ranges of the PDSMF, this value is only a lower limit to the mass of the cluster. As mentioned in Sect. 4.6.1, the uncertainties in the mass-luminosity relations are yet to be established. Thus the quoted uncertainties of our mass results are underestimated.

Figure 4.13: Normalised PDSMF in J, H, K_s bands. Also shown the IMFs of Chabrier (2005); Thies & Kroupa (2007) and fits to the PDSMF found by us and Bouy et al. (2015b).

.

4.7 The mass distribution on time

In order to test if dynamical effects could be the origin of the discrepancies between the PDSMFs and the IMFs, in this section I analyse the evolution in time of the mass distribution. To do this, I compare the differences between the Pleiades PDSMF (≈ 120 Myr) and those of the Trapezium and Hyades clusters, which are ≈ 1 Myr and ≈ 600 Myr old respectively. These can be thought as snapshots of the Pleiades pasts and future mass functions.

In Figure 4.14 I compare the PDSMF from the Pleiades derived here, to those of the Trapezium and Hyades. These PDSMFs correspond to those of Fig. 11 of Bouy et al. (2015b) (private communication). As mentioned by Bouy et al. (2015b), the abundance of low-mass stars and brown dwarfs in the range $0.03 - 0.1$ $M_\odot (\log M/M_\odot \approx \{-1, -1.4\})$ seems to diminish with time. Furthermore, since the PDSMF are normalised, a diminish in the lower mass range produces a relative increase of low-mass stars in the range $-0.4 < \log M/M_\odot < -0.2$. The alternative effect of an increase in the high mass range which then may produce a relative diminish in the low mass range, although is statistically possible, is unlikely in the astrophysical sense. Open clusters tend lose stars, not the other way around. Furthermore, this effect is consistent with the classical scenario in which low-mass stars and brown dwarfs are ejected as the cluster relaxes. I test the validity of this scenario, at least the statistical significance of the observed differences among the PDSMF of this three clusters. To perform this test, the null hypothesis is that the Trapezium and the Hyades have the same PDSMF as the Pleiades. Since I only have the inferred model of the Pleiades cluster, I am compelled to perform a frequentist test. Thus, to do the statistical comparison of these three PDSMF I use the Kolmogorov-Smirnov and the Anderson-Darling tests.

Figure 4.15 shows the cumulative distribution functions (CDFs) of the Trapezium, Pleiades (only in K_s band) and Hyades PDSMFs. Also and for comparison,

Figure 4.14: Left: PDSMFs of the Pleiades (derived here for J, H, K_s bands), Trapezium, and Hyades, from Bouy et al. (2015b). They are normalised in the interval of completeness.

.

Figure 4.15: Cumulative distribution functions (CDF) of the PDSMFs from left panel and that of Chabrier (2005) and Thies & Kroupa (2007) system initial mass function (normalised also in the interval of completeness). The Pleiades CDF shown is just from K_s band. The grey area depicts the area in which the null hypothesis of same PDSMF as that of the Pleiades can not be rejected (at $\alpha = 0.01$).

.

we show the CDFs of Chabrier (2005) and Thies & Kroupa (2007) IMFs. The grey area around the Pleiades CDF shows the hypothesis test in which I compare each CDF with that of the Pleiades. The null hypothesis,as mentioned before is that each compared CDF is exactly that of the Pleiades. I use the Kolmogorov-Smirnov statistic and the alpha value $\alpha = 0.01$, to compute the maximum vertical distance d_α from the Pleiades CDF, the grey region was created with this maximum distance. The null hypothesis is rejected only if the tested CDF lies entirely outside the grey region around the Pleiades CDF. As can be seen, neither the IMFs nor the PDSMF of the Trapezium and Hyades lay entirely within the grey area, thus we can reject the null hypothesis that they share the same PDSMF of the Pleiades. Furthermore, since the Kolmogorov-Smirnov test uses only the maximum distance between CDFs, we also applied the more robust Anderson-Darling test. It also rejects the null hypotheses (at $p < 0.004$) that the Trapezium and Hyades PDSMFs and the Chabrier (2005) and Thies & Kroupa (2007) IMFs have the same CDF of the Pleiades.

The previous tests show that there is enough evidence to claim for differences among the PDSMFs of these three clusters and from IMFs and Pleiades PDSMF. Thus they suggest that the observed differences may have an origin on dynamical effects associated with age and relaxation. Nevertheless, to claim for reliable evidence supporting these differences the census of the Trapezium and Hyades must be done using the same methods. Also, the uncertainties must be properly estab-

lish both for the other PDSMFs and for the mass-luminosity relation from which all these PDSMF are derived.

Chapter 5

Conclusions and Future Work

Bibliography

????

Adams, J. D., Stauffer, J. R., Monet, D. G., Skrutskie, M. F., & Beichman, C. A. 2001a, AJ, 121, 2053

Adams, J. D., Stauffer, J. R., Monet, D. G., Skrutskie, M. F., & Beichman, C. A. 2001b, AJ, 121, 2053

Adams, W. S. 1904, ApJ, 19

Akeret, J., Seehars, S., Amara, A., Refregier, A., & Csillaghy, A. 2013, Astronomy and Computing, 2, 27

Allard, F. 2014, in IAU Symposium, Vol. 299, Exploring the Formation and Evolution of Planetary Systems, ed. M. Booth, B. C. Matthews, & J. R. Graham, 271–272

Allard, F., Homeier, D., & Freytag, B. 2013, Mem. Soc. Astron. Italiana, 84, 1053

Allard, F., Homeier, D., Freytag, B., & Sharp, C. M. 2012, in EAS Publications Series, Vol. 57, EAS Publications Series, ed. C. Reylé, C. Charbonnel, & M. Schultheis, 3–43

Alsing, J., Heavens, A., & Jaffe, A. H. 2017, MNRAS, 466, 3272

Anderes, E., Wandelt, B. D., & Lavaux, G. 2015, ApJ, 808, 152

Andrieu, C., de Freitas, N., Doucet, A., & Jordan, M. I. 2003, Machine Learning, 50, 5

Astraatmadja, T. L. & Bailer-Jones, C. A. L. 2016a, ApJ, 832, 137

Astraatmadja, T. L. & Bailer-Jones, C. A. L. 2016b, ApJ, 833, 119

Azzalini, A. 1996, Biometrika, 83, 715

Azzalini, A. & Capitanio, A. 1999, Journal of the Royal Statistical Society: Series B (Methodological)

Bailer-Jones, C. A. L. 2015, PASP, 127, 994

Balaguer-Núñez, L., Galadí-Enríquez, D., & Jordi, C. 2007, A&A, 470, 585

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403

Barnes, III, T. G., Moffett, T. J., Jefferys, W. H., & Forestell, A. D. 2004, in Astronomical Society of the Pacific Conference Series, Vol. 310, IAU Colloq. 193: Variable Stars in the Local Group, ed. D. W. Kurtz & K. R. Pollard, 95

Barrado, D., Bouy, H., Bouvier, J., et al. 2016, A&A, 596, A113

Bishop, C. 2006, Pattern Recognition and Machine Learning, Information Science and Statistics (Springer)

Blaauw, A. 1964, in IAU Symposium, Vol. 20, The Galaxy and the Magellanic Clouds, ed. F. J. Kerr, 50

Blackwell, T. & Bentley, P. 2002, in Proceedeings of the 4th Annual Conference on Genetic and Evolutionary Computations, ed. W. Langdon, E. Cantu-Paz, K. Mathias, R. Roy, & D. Davis (San Francisco, CA, USA: Morgan kaufmann Publishers Inc.), 19–26

Bouvier, J., Kendall, T., Meeus, G., et al. 2008, A&A, 481, 661

Bouy, H., Bertin, E., Barrado, D., et al. 2015a, A&A, 575, A120

Bouy, H., Bertin, E., Moraux, E., et al. 2013, A&A, 554, A101

Bouy, H., Bertin, E., Sarro, L. M., et al. 2015b, A&A, 577, A148

Bovy, J., Hogg, D. W., & Roweis, S. T. 2011, The Annals of Applied Statistics, 5, 1657

Brooks, S., Gelman, A., Jones, G., & Meng, X. 2011, Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC Handbooks of Modern Statistical Methods (CRC Press)

Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Chabrier, G. 2003, PASP, 115, 763

Chabrier, G. 2005, in Astrophysics and Space Science Library, Vol. 327, The Initial Mass Function 50 Years Later, ed. E. Corbelli, F. Palla, & H. Zinnecker, 41

Chung, Y., Gelman, A., Rabe-Hesketh, S., Liu, J., & Dorie, V. 2015, Journal of Educational and Behavioral Statistics, 40, 136

Clerc, M. & Kennedy, J. 2002, Trans. Evol. Comp, 6, 58

Converse, J. M. & Stahler, S. W. 2008, ApJ, 678, 431

Converse, J. M. & Stahler, S. W. 2010, MNRAS, 405, 666

D'Antona, F. 1998, in Astronomical Society of the Pacific Conference Series, Vol. 142, The Stellar Initial Mass Function (38th Herstmonceux Conference), ed. G. Gilmore & D. Howell, 157

De Boor, C. 1978, A practical guide to splines / Carl de Boor (Springer-Verlag New York), xxiv, 392 p. :

Demory, B.-O. 2014, ApJ, 789, L20

Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B, 39, 1

Diard, J. & Bessiere, P. 2008, Probabilistic reasoning and decision making in sensory-motor systems, 153

Dirac, P. A. M. 1942, Proceedings of the Royal Society of London Series A, 180, 1

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. 1987, Physics Letters B, 195, 216

Duquennoy, A. & Mayor, M. 1991, A&A, 248, 485

Faherty, J. K., Burgasser, A. J., Walter, F. M., et al. 2012, ApJ, 752, 56

Fan, Y. & Sisson, S. 2011, Handbook of Markov Chain Monte Carlo, 67

Feeney, S. M., Johnson, M. C., McEwen, J. D., Mortlock, D. J., & Peiris, H. V. 2013, Phys. Rev. D, 88, 043012

Fei-Fei, L. & Perona, P. 2005, in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 2, Ieee, 524–531

Feroz, F., Hobson, M. P., & Bridges, M. 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601

Flegal, J. M., Hughes, J., & Vats, D. 2016, mcmcse: Monte Carlo Standard Errors for MCMC, Riverside, CA and Minneapolis, MN, r package version 1.2-1

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publications of the Astronomical Society of the Pacific, 125, 306

Fraley, C. & Raftery, A. E. 2002, Journal of the American Statistical Association, 97, 611

Fraley, C., Raftery, A. E., Murphy, T. B., & Scrucca, L. 2012, mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation

Franklin, J. 2002, The Science of Conjecture: Evidence and Probability Before Pascal, A Johns Hopkins paperback (Johns Hopkins University Press)

Gagné, J., Lafrenière, D., Doyon, R., Malo, L., & Artigau, É. 2014, ApJ, 783, 121

Gaia Collaboration, van Leeuwen, F., Vallenari, A., et al. 2017, A&A, 601, A19

Galli, P. A. B., Moraux, E., Bouy, H., et al. 2017, A&A, 598, A48

Gatewood, G., de Jonge, J. K., & Han, I. 2000, ApJ, 533, 938

Gelman, A. 2006, Bayesian Analysis, 1, 515

Gelman, A. 2012, CHANCE, 25, 52

Gelman, A., Carlin, J., Stern, H., et al. 2013, Bayesian Data Analysis, third edition edn. (CRC)

Gelman, A. & Hill, J. 2007, Data Analysis Using Regression and Multilevel/Hierarchical Models, Analytical Methods for Social Research (Cambridge University Press)

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. 2008, Ann. Appl. Stat., 2, 1360

Geman, S. & Geman, D. 1984, IEEE Trans. Pattern Anal. Mach. Intell., 6, 721

Gong, L. & Flegal, J. M. 2016, Journal of Computational and Graphical Statistics, 25, 684

González-Farias, G., Dominguez-Molina, A., & Gupta, A. K. 2004, Journal of Statistical Planning and Inference, 126, 521

Good, I. J. 1980, Trabajos de Estadistica Y de Investigacion Operativa, 31, 489

Goodman, J. & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Gupta, A. K., González-Farias, G., & Dominguez-Molina, J. 2004, Journal of Multivariate Analysis, 89, 181

Guthrie, B. N. G. 1987, QJRAS, 28, 289

Hambly, N. C. & Jameson, R. F. 1991, MNRAS, 249, 137

Hand, D. J. & Yu, K. 2001, International Statistical Review / Revue Internationale de Statistique, 69, 385

Hastings, W. K. 1970, Biometrika, 57, 97

Hertzsprung, E. 1947, Annalen van de Sterrewacht te Leiden, 19, A1

Hogg, D. W., Bovy, J., & Lang, D. 2010a, ArXiv e-prints

Hogg, D. W., Myers, A. D., & Bovy, J. 2010b, ApJ, 725, 2166

Hong, D., Balzano, L., & Fessler, J. A. 2016, Towards a theoretical analysis of PCA for heteroscedastic data, arxiv 1610.03595

Huang, A. & Wand, M. P. 2013, Bayesian Analysis, 8, 439

Jaynes, E. 2003, Probability Theory: The Logic of Science (Cambridge University Press)

JCGM. 2008, JCGM 100: Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement, Tech. rep., Joint Committee for Guides in Metrology

Jefferys, T. R., Jefferys, W. H., Barnes, III, T. G., & Dambis, A. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 371, Statistical Challenges in Modern Astronomy IV, ed. G. J. Babu & E. D. Feigelson, 433

Johnson, H. L. & Mitchell, R. I. 1958, ApJ, 128, 31

Jones, B. F. 1970, AJ, 75, 563

Kass, R. E. & Raftery, A. E. 1995, Journal of the American Statistical Association, 90, 773

Kennedy, J. & Eberhart, R. 1995, in Neural Networks, 1995. Proceedings., IEEE International Conference on, Vol. 4, 1942–1948 vol.4

King, I. 1962, AJ, 67, 471

Koller, D. & Friedman, N. 2009, Probabilistic Graphical Models: Principles and Techniques, Adaptive computation and machine learning (MIT Press)

Krone-Martins, A. & Moitinho, A. 2014, A&A, 561, A57

Lee, S.-W. & Sung, H. 1995, Journal of Korean Astronomical Society, 28, 45

Limber, D. N. 1961, ApJ, 134, 537

Limber, I. D. N. 1962, ApJ, 135, 16

Lin, T., Lee, J., & Ho, H. 2006, WOS:000236696400015

Lindstrom, M. 1999, Journal of Computational and Graphical Statistics, 8, 333

Liu, Z., Chen, Y., Tian, S., & Xu, Z. 2015, Journal of Information and Computational Science, 12, 775

Lodieu, N., Deacon, N. R., & Hambly, N. C. 2012, MNRAS, 422, 1495

Loktin, A. V. 2006, Astronomy Reports, 50, 714

Lutz, T. E. & Kelker, D. H. 1973, PASP, 85, 573

Malo, L., Doyon, R., Lafrenière, D., et al. 2013, ApJ, 762, 88

March, M. C., Karpenka, N. V., Feroz, F., & Hobson, M. P. 2014, MNRAS, 437, 3298

McMichael, D. W. 1996, in in Proc. Fourth International Symposium on Signal Processing and its Applications (ISSPA, 377–378

Melis, C., Reid, M. J., Mioduszewski, A. J., Stauffer, J. R., & Bower, G. C. 2014, Science, 345, 1029

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, The Journal of Chemical Physics, 21, 1087

Moraux, E., Bouvier, J., & Stauffer, J. R. 2001, A&A, 367, 211

Morris, C. N. 1983, Journal of the American Statistical Association, 78, 47

Muench, A. A., Lada, E. A., Lada, C. J., & Alves, J. 2002, ApJ, 573, 366

Neal, R. M. 1996, Bayesian Learning for Neural Networks (Secaucus, NJ, USA: Springer-Verlag New York, Inc.)

Olivares, J. & et al. 2017, in preparation

Olivares, J., Sarro, L., Moraux, E., Berihuete, A., & Bouy, H. 2017, submitted to å

Ozerov, A., Lagrange, M., & Vincent, E. 2013, Computer Speech and Language, 27, 874 , special Issue on Speech Separation and Recognition in Multisource Environments

Patel, P. K., Sharma, V., & Gupta, K. 2013, International Journal of Computer Applications, 73

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49

Pinfield, D. J., Jameson, R. F., & Hodgkin, S. T. 1998, MNRAS, 299, 955

Press, W. H. 1997, in Unsolved Problems in Astrophysics, ed. J. N. Bahcall & J. P. Ostriker, 49–60

Pritchard, C. 1884, MNRAS, 44, 355

Pugachev, V. 1965, Theory of Random Functions and its Application to Control Problems (Pergamon), 852

Raboud, D. & Mermilliod, J.-C. 1998a, A&A, 329, 101

Raboud, D. & Mermilliod, J.-C. 1998b, A&A, 329, 101

Riedel, A. R., Blunt, S. C., Lambrides, E. L., et al. 2017, AJ, 153, 95

Roberts, G. O. & Rosenthal, J. S. 2004, Probability Surveys, 1, 20

Sale, S. E. 2012, MNRAS, 427, 2119

Sampedro, L. & Alfaro, E. J. 2016, MNRAS, 457, 3949

Sarro, L. M., Bouy, H., Berihuete, A., et al. 2014, Astronomy & Astrophysics, 14

Schwarz, G. 1978, Ann. Statist., 6, 461

Shariff, H., Dhawan, S., Jiao, X., et al. 2016, MNRAS, 463, 4311

Shkedy, Z., Decin, L., Molenberghs, G., & Aerts, C. 2007, MNRAS, 377, 120

Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, American Institute of Physics Conference Series, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405

Skilling, J. 2006, Bayesian Anal., 1, 833

Smyth, G. K. 1998, in Encyclopedia of Biostatistics, ed. Armitage, P. & Colton, T. (John Wiley and Sons)

Soderblom, D. R., Nelan, E., Benedict, G. F., et al. 2005, AJ, 129, 1616

Spiriti, S., Eubank, R., Smith, P. W., & Young, D. 2013, Journal of Statistical Computation and Simulation, 83, 1020

Stauffer, J. R., Hartmann, L. W., Fazio, G. G., et al. 2007, ApJS, 172, 663

Stauffer, J. R., Schultz, G., & Kirkpatrick, J. D. 1998, ApJ, 499, L199

Stoneking, C. J. 2014

Takeda, Y., Hashimoto, O., & Honda, S. 2017, PASJ, 69, 1

Tapiador, D., Berihuete, A., Sarro, L. M., Julbe, F., & Huedo, E. 2017, Astronomy and Computing, 19, 1

Thies, I. & Kroupa, P. 2007, ApJ, 671, 767

Titus, J. 1938, AJ, 47, 25

Trumpler, R. J. 1921, Lick Observatory Bulletin, 10, 110

van Leeuwen, F. 1980, in IAU Symposium, Vol. 85, Star Clusters, ed. J. E. Hesser, 157–162

van Leeuwen, F. 1999, A&A, 341, L71

van Leeuwen, F. 2009, A&A, 497, 209

Wolfgang, A. & Lopez, E. 2015, ApJ, 806, 183

Woolley, R. V. D. R. 1956, MNRAS, 116, 296